Thesis for the Master's degree in Molecular Biosciences

Development of Multi Locus Variable Number Tandem Repeat Analysis

(MLVA) for the genotyping of *Legionella pneumophila* isolated from various habitats.

Anusha Balasingham

60 study points

Department of Molecular Biosciences

Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO 03/2008

Division of Environmental Medicine,

NORWEGIAN INSTITUTE OF PUBLIC HEALTH

ACKNOWLEDGEMENTS

This project was carried out in the two of the Norway's largest institutions in research, the National Public Health Institute and University of Oslo, Norway during 2007-2008.

I wish to express my warm thanks to the professors who introduce this project to me and provided the excellent facilities for this study.

My genuine thanks and respect belong to my internal supervisor Reidun Sirevåg, University of Oslo and external supervisor Hans Utkilen, National Public Health Institute, Oslo, Norway.

I wish to thank to Professor Kjetill S Jakobsen for being actively engaged and supporting me during this project. I really admire your positive attitude.

Thanks to Post-doctors, Mary J. Anderson-Glenna and Alexander J. Nedberget, who gave me scientific guidance, excellent cooperation, and friendly support during this period. I thank them for always taking time in their hectic schedule to answering my endless questions and being patient with me and for encouraging and helping me in the process of writing the thesis. It was a pleasure to work with you. I wish you good luck for your future research.

I thank also to all the members at the Department of MIVA, The National Public Health Institute and CEES Lab, University of Oslo for providing such a nice working place and sharing good company and scientific input.

Finally I thank my family. I am very grateful for your cooperation and support. Without your support I could not had enough time to complete my degree.

ABSTRACT

Legionella is present worldwide and is a part of natural aquatic habits. *Legionella* is also a pathogen which is the causative agent of Legionnaires disease. When incidents of Legionnaires disease are identified, there is a need not only to identify the causative agent but to trace the source and prevent further distribution.

In this study we adapted an existing high resolution genetic typing method known as MLVA from agarose gel based to capillary electrophoresis based analysis. MLVA is based on minisatellite analysis.

Subsequently this method was applied to a number of environmentally and clinically isolated *L. pneumophila* strains. To the best of our knowledge this is the first time that capillary electrophoresis has been used as part of the MLVA analysis of *L. pneumophila*.

Capillary electrophoresis is a very sensitive and robust technique; in order to adapt the method for capillary electrophoresis it was necessary to optimise the polymerase chain reaction (PCR). PCR optimization was performed by modifying reaction conditions including the annealing temperature, concentration of magnesium chloride, PCR reaction volume and choice of DNA polymerase enzyme.

The development of capillary electrophoresis has several advantages; such as speed, high separation efficiency, low sample consumption and the ability to analyze multiple PCR products in the same capillary. By using CE analysis two additional alleles were observed for one of the markers when analysing strains within the proficiency panel.

All together 27 *L. pneumophila* were analysed by MLVA typing and 14 different genotypes of *L. pneumophila* were observed. These data can be maintained as a genotype library in Norwegian Institute of Public Health. Future outbreaks or clinical strains can be compared to the growing database of MLVA genotypes.

It is hoped that this technique will be used during future studies which aim to track *L. pneumophila* during outbreak situations.

ABBREVIATIONS

ABI-Applied Biosystems
AE-Agarose gel electrophoresis
Bp-base pairs
CE -Capillary Electrophoresis
DA- difficult to analysis.
DNA- deoxyribonucleic acid
dH_2O- distilled water
dNTP-Deoxyribonucleotide triphosphate
Etbr-Ethidium bromide
Fig- Figure
EWGLI-European Working Group for *Legionella* Infections
Kb- kilobasepair(s)
Ldr-Ladder
Lpms- *Legionella pneumophila* minisatellites
L.pnemophila- Legionella pneumophila
MIP - Macrophage infectivity potentiator protein
MLVA - multilocus variable number tandem repeat analysis
MLST-Multilocus sequence typing
NCBI-National Center for Biotechnology
NA- No amplification Information
N.g: New genotype
NCTC-National collection of type cultures
PCR - Polymerase Chain Reaction
S- second(s)
T-temperature in ^oC.
TD-touchdown PCR
UIO-University in Oslo
U-unit
VNTR - variable number tandem repeats

CONTENTS

Preface

Part of this dissertation has been accepted for publication, in the Journal of Microbiological Methods (2008), doi:10.1016/j.mimet.2008.02.007

Title- Multiple-locus variable-number tandem repeat analysis of *Legionella pneumophila* using multi-colored capillary electrophoresis.

Authors- Nederbragt, Alexander J., Balasingham, Anusha, Sirevåg,

Reidun, Utkilen, Hans, Jakobsen, Kjetill S., Anderson-Glenna, Mary J.

This article is available online at:

http://dx.doi.org/10.1016/j.mimet.2008.02.007

1.Introduction

1.1. *Legionella* bacteria

Legionella was first recognized in association with an outbreak of pneumonia that occurred among attendees of an American Legion convention in Philadelphia in 1976 (Fraser, Tsai et al. 1977; McDade, Shepard et al. 1977). The disease was dubbed Legionnaire`s disease by the media. Dr.Joseph McDade discovered *Legionella pneumophila* as the etiological agent of the outbreak, naming the bacterium after this outbreak (http://gsbs.utmb.edu/microbook/ch040.htm). Subsequently many different species of *Legionella,* have been isolated from the natural environment and from clinical samples. The number of identified cases of Legionnaires disease has been on the rise in recent years (Wullings and van der Kooij 2006). It is difficult to ascribe the number of cases of Legionnaires disease but it has been estimated that in the USA 20 cases per million people occur each year and up to 34 cases per million occur in Europe. In Europe 10,322 cases of Legionnaires disease were reported for a 2 year period from 2000 (Joseph 2004) In the case of non outbreak situations the increase in reported cases may be related to improved detection methods or better reporting. Recently, in May 2005, a fatal outbreak of Legionnaires disease occurred in the Sarpsborg/Fredrikstad area of Norway. 52 people were infected with 10 deaths (Nygard 2005). Eventually the source of the infective *L. pneumophila* was traced to an air scrubber of a local industrial installation. This outbreak within our own country highlights the need for further research in all areas of

Legionellae biology and ecology. Figure 1A. Shows electron microscopy image of the *L. pneumophila*.

Figure .1A. Electron microscopy images of *Legionella pneumophila* http://www.denniskunkel.com/product_info.php?products_id=439

1.2. Members of the family Legionellaceae

The genus *Legionella* belongs to the phylum Proteobacteria, the class Gammaproteobacteria, the order *Legionellales* and family *Legionellaceae* (Brenner, Krieg et al. 2005). The *Legionellaceae* contains approximately 50 separate species which can be further subdivided into subspecies and serogroups (Stolhaug and Bergh 2006). *L. pneumophila* is the most studied and most important member of the family with illnesses being ascribed to it in over 90% of cases (Aurell, Farge et al. 2005). *L. pneumophila* has 3 subspecies, *pneumophila*, *fraseri* and *pascullei* and 15 serotypes, with serogroup 1 causing the majority of outbreaks (Helbig et al., 2002). A list of currently known *Legionella* bacterial species is listed in Table 1.

Table .1. A list of currently known *Legionella* bacterial species. (Fields, Benson et al. 2002).

No	species	No. Of serogroups	associated with disease or not
1	*L.pneumophila*	15	Causative agent of human pneumonia, empyema and Pontiac fever
2	*L. adelaidensis-*	1	Not demonstrated to cause human disease- Isolated from cooling tower water in Adelaide, Australia in 1987.
3	*L. anisa*	1	Causative agent of human pneumonia, empyema and Pontiac fever - Isolated in Los Angeles, California, from hot water in a sink by George W. Gorman in 1981.
4	*L. birminghamensis*	1	Causative agent of human pneumonia -Isolated from a human lung biopsy in Birmingham, Alabama in 1986.
5	*L. bozemanii*	2	Causative agent of human pneumonia and empyema. Isolated from human lung tissue in 1959 by Bozeman.
6	*L. brunensis*	1	No evidence for human disease. Isolated from cooling tower water in brno, Czechoslovakia; frist reporded in 1986.
7	*L. cherrii*	1	Not demonstrated to cause human disease - Isolated from thermally polluted water in Minnesota by R.L. Tyndall and C.B. Duncan in 1982.
8	*L.cincinnatiensis*	1	Causative agent of human pneumonia **-** Isolated from a human open lung biopsy in Cincinnati, Ohio in 1982
9	*L. dumoffii*	1	Causative agent of human pneumonia **-** Isolated from water by G. W. Gorman in 1977.
10	*L. erythra*	2	Not yet demonstrated to cause human disease - Isolated from cooling tower water in seattle, Washington, by George W. Gorman in 1981.
11	*L. fairfieldensis*	1	Not yet demonstrated to cause human disease - Isolated from cooling tower water in Fairfield, Victoria Australia, in 1987.
12	*L. feeleii*	2	Causative agent of human pneumonia and acute, nonpnemonic, febrile illness (pontanic fever) **-** Isolated from industrial coolant by George W. Gorman in 1981.
13	*L. geestiana*	1	Not demonstrated to cause human disease - Isolated from a domestic hot water tap in the Geest Office Building in London, United Kingdom, between 1982 and 1984.
14	*L. gormanii*	1	Causative agent of human pneumonia **-**Isolated by G.W. Gorman creek bank soil in 1978.
15	*L. gratiana*	1	Not demonstrated to cause human disease - Isolated from spring water at a spa in the savoy region of France, frist repordetd in 1989.
16	*L. hackeliae*	2	Causative agent of human pneumonia **-** Isolated from a bronchial biopsy specimen by Meredith Hackel in 1981.
17	*L. israelensis*	1	Not demonstrated to cause human disease - Isolated from an oxidation pond in gaash, Israel, Herve Bercovier.
18	*L.jamestowniensis*	1	Not demonstrated to cause human disease - Isolated by G.W. Gorman

			from wet soil in 1979.
19	*L. jordanis*	1	Causative agent of human pneumonia - Isolated from surface water by G.W. Gorman in 1978.
20	*L. lansingensis*	1	Causative agent of human pneumonia - Isolated from a human bornchoscopic washing in Lansing, Michigan, in 1986.
21	*L. londiniensis*	1	Not demonstrated to cause human disease - Isolated between 1982 and 1984 from cooling tower pond water in London, England.
22	*L. longbeachae*	2	Causative agent of human pneumonia - Isolated from a transtracheal aspirate by R. Porschen in 1980.
23	*L. lytika*	1	Amoebal pathogen, etiologic agent of human disease - Isolated from human sputum and lung with amoebae.
24	*L. maceachernii*	1	Causative agent of human pneumonia - Isolated by George W. Gorman from a home evaporative condenser in 1979.
25	*L. micdadei*	1	Causative agent of human pneumonia - Isolated by H. Tatlock from a guinea pig inoculated with human blood in 1943.
26	*L. moravica*	1	Not demonstrated to cause human disease - Isolated from cooling tower water in Moravia, where the species was first Isolated; first reported in1987.
27	*L.nantarum*	1	Not demonstrated to cause human disease - Isolated from a domestic water system in London, England.
28	*L. oakridgensis*	1	Causative agent of human pneumonia -Isolated from cooling tower water by R.L Tyndall, C.B Duncan, and E.L Domingue in 1981.
29	*L. parsisensis*	1	Causative agent of human pneumonia - Isolated from cooling tower water by George W. Gorman in 1981.
30	*L. quatteirensis*	1	Not demonstrated to cause human disease - Isolated from a shower in a bathroom of a hotel in Quateria, Portugal, between 1982 and 1984.
31	*L.quinlivanii*	2	Not demonstrated to cause human disease - Isolated from water in Australia in 1986.
32	*L. rubrilucens*	1	Not demonstrated to cause human disease - Isolated from tap water in Los Angeles, California, by Giorge W. Gorman in 1980.
33	*L. sainthelensi*	2	Causative agent of human pneumonia - Isolated from spring water by J. Campbell and S. Eng in 1981.
34	*L. santicrucis*	1	Not demonstrated to cause human disease - Isolated from tap water in St. Croix, Virgin Islands, by George W. Gorman in 1982.
35	*L. shakespearei*	1	Not demonstrated to cause human disease - Isolated from cooling tower water in Stratford-upon-Avon, United Kingdom, reported in 1992.
36	*L. spiritensis*	1	Not yet associate with human disease. There are two serogroups - Isolated from lake water by J.Campbell in 1981.
37	*L. steigerwaltii*	1	Not demonstrated to cause human disease - Isolated from tap water in St. Croix, Virgin Islands, by George W. Gorman in 1982.
38	*L. taurinensis*	1	Not demonstrated to cause human disease - Isolated from a humidifider in Turin, Italy.

39	*L.tucsonensis*	1	Causative agent of human pneumonia -Isolated from human pleural fluid in Tucson, Arizona in 1984.
40	*L.wadsworthii*	1	Causative agent of human pneumonia - Isolated from sputum by Paul H. Edelstein in 1981.
41	*Legionella waltersii*	1	Not demonstrated to cause human disease - Isolated from a drinking water distribution system in South-Australia
42	*L. worsleiensis*	1	Not demonstrated to cause human disease - Isolated from cooling tower return flow at an industrail site in Worsley United Kingdom, between 1982 and 1984.
43	*L.genomospecies*	1	Not demonstrated to cause human disease
44	L.drozanskii	1	Not demonstrated to cause human disease
45	*L. rowbothamii*	1	Not demonstrated to cause human disease
46	*L.fallonii*	1	Not demonstrated to cause human disease
47	*L.gresilensis*	1	Not demonstrated to cause human disease
48	*L.beliardensis*	1	Not demonstrated to cause human disease

1.3. Characteristics of the family Legionellaceae

The *Legionellaceae* are aerobic chemoorganotrophs. They use amino acids as sources of carbon and energy. L-cysteine-HCl and iron salts are essential requirements for growth. The bacteria are urease and oxidase negative. Morphologically they appear as Gram negative non-capsulated rods or filaments, 0.3-0.9 x 2-20 µm in length. Figure 1B. Shows electron microscopy image of the *L. pneumophila.* The majority of *Legionellaceae* are motile and have one or more curved or straight flagellae with polar or lateral arrangements (Brenner, Krieg et al. 2005).

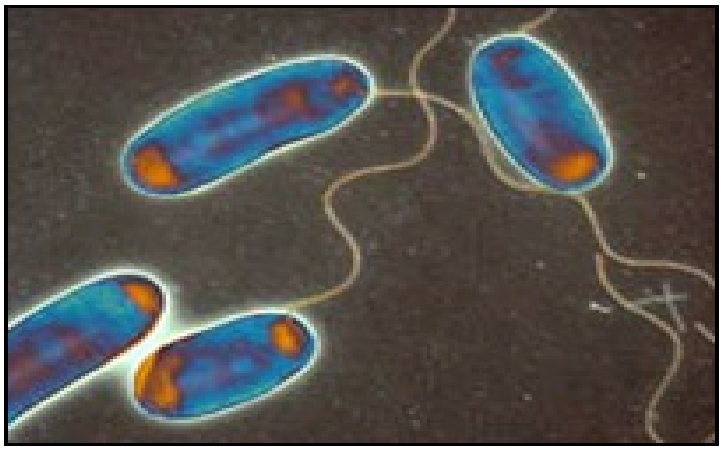

Figure .1B. Electron microscopy images of *Legionella pneumophila* http://www.okkjemi.no/Legionella/legionella2.jpg

1.4. Ecology

Members of the genus *Legionella* are wide spread in natural aquatic systems such as rivers and lakes, albeit at low concentrations typically less than 1% of the bacterial population (Fliermans, Cherry et al. 1981). Within manmade installations including cooling towers, air conditioning systems and water distribution systems they multiply to greater levels. Temperature is an important factor affecting the occurrence and multiplication of *Legionella* in the aquatic environment. The bacteria have been found in water temperatures ranging from 6-60^{o}C, but do not proliferate below 20^{o}C and die above 60^{o}C. Optimum temperatures for growth lie between 30 ^{o}C and 40^{o}C. Interestingly the bacteria's cellular fatty acid composition is similar to that of thermophilic bacteria (Moss, Weaver et al. 1977). It has also been noted that virulence varies with temperature, *L. pneumophila* strains grown at 37^{o}C are more virulent than those incubated at 24^{o}C (Mauchline, James et al. 1994).

pH of the aquatic environment has an influence on the growth of *Legionella* with detection in waters with pHs between 5.5 and 8.1(Fliermans, Cherry et al. 1981). Laboratory tests revealed that multiplication of the bacteria could occur between pH 5.5 and 9.2 (Wadowsky, Wolford et al. 1985). The presence of high iron concentrations in water distribution systems has been correlated with the growth of *Legionella* (States, Conley *et al.* 1985). *Legionella* have been shown to survive in biofilms. The presence of *Legionella* in the environment is also influenced by the presence of a number of free-living amoeba and ciliated protozoa which serve as a host (Fields, Benson et al. 2002; Steinert, Hentschel et al. 2002).

1.5. Intracellular life within single celled organisms

It is generally accepted that in the natural environment protozoan species are essential for the growth of *Legionella*. Ecological studies have shown that protozoa serve as hosts for *Legionella* in a variety of environments where they replicate within the protozoan vacuoles for extended periods eventually killing the host (Fields, Benson et al. 2002). The ability of *L.pneumophila* to multiply intracellularly in protozoa was first described by Rowbotham in 1980 (Rowbotham 1980). *Legionella* are known to multiply in 14 species of amoeba, two species of ciliated protozoa, and one species of slime mold (Fields, Benson et al. 2002). Amoebae are phagocytes; normally they engulf bacterial cells as food sources and are thought to play a role in controlling bacterial densities in the environment. Protozoa also serve as shelter for

Legionella during harsh conditions such as drying, high temperatures, variations in osmolarity, pH variations, chlorination and exposure to biocidal compounds. Figure 2. shows how they infect an amoeba (its environmental host) and replicate inside it.

It has been observed that *Legionella* may avoid disinfection during water treatment in this way. Invasion and intracellular replication of *L. pneumophila* within protozoa in the environment plays a major role in the transmission of Legionnaires' disease (Atlas 1999).

Figure.2. Legionella bacteria infection of amoeba and replication inside their environmental hosts.

1.6. *Legionella* within biofilms

Although evidence suggests that protozoa are a key factor in *Legionella* multiplication process and survival in nature, it is known that biofilms also support their survival and proliferation (Atlas 1999). Biofilms are complex microbial communities attached at interfaces and bound with a matrix of extracellular polymeric substances (O'Toole, Kaplan et al. 2000). The natural mode of growth of bacteria in the environement is within biofilms. Biofilm formation occurs in a number of stages (figure 3). Bacteria and other micro-

organisms (such as protozoa, fungi and algae) attach to surfaces first reversibly and afterwards irreversibly bound with the help of a sticky polymeric matrix. Existence within biofilms improves survival and growth. Biofilms provide protection to micro-organisms members from biocides and predation as well as acting as a nutrient hot spot. There is evidence that majority of out breaks of *Legionella* are associated with biofilms in water distribution systems (Fields, Benson et al. 2002). Additionally biofilms within hospital water distribution systems pose a serious threat to already weak individuals.

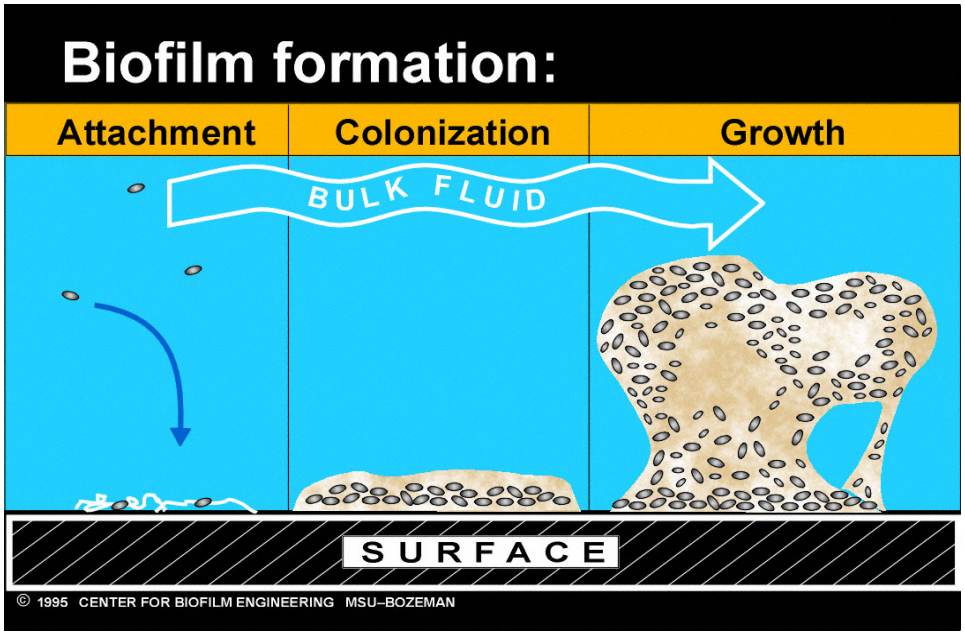

Figure.3. Biofilm formation by microorganisms
http://www.uweb.engr.washington.edu/images/research/biofilmtutorial.JPG

1.7 Disease caused by *Legionella*.

Legionella are intracellular parasites of free-living amoebae and protozoa and use the same mechanism to multiply within mammalian macrophages cells. The two main types of legionellosis caused by *Legionella* are

Legionnaires disease and Pontiac Fever. Evidence suggests that *L.pneumophila* is the most significant species and is predominantly associated with outbreaks of legionnaire's disease. The disease itself is caused when critical numbers of *Legionella* in the aquatic environment become aerosolized, i.e. form tiny droplets of water in the air, and are inhaled. (Jantzen and Olsen 2002) reported that densities above 10^4 to 10^5 CFU/liter represent a possible threat to human health. (Joly, Falconnet et al. 2006) *Legionella* do not infect humans through drinking of contaminated water or from person to person contact (Jantzen and Olsen 2002).

1.7.1. Legionnaire's disease

Legionnaire's disease is an acute respiratory illness of variable severity with a significant mortality rate of about 15 to 20% (Joly, Falconnet et al. 2006). The incubation time is 2 to 10 days. Symptoms include shivering and coldness with a rise in body temperature, a non productive cough and difficulty breathing. Figure 4. Shows the human lung infected by legionnaire`s disease. The disease may occur in all age groups of healthy individuals but those at highest risk have been shown to be males over the age of 55 (Pattison, N. et al. 2006). Additionally smokers, those with other lung diseases, diabeticses or immunocompromised individuals are all at increased risk. The disease is usually treated with the antibiotic erythromycine but erythromycin-resistant strains have also emerged Alternatively a combination with rifampicin or doxycycline can be used. (Brenner, Krieg et al. 2005).

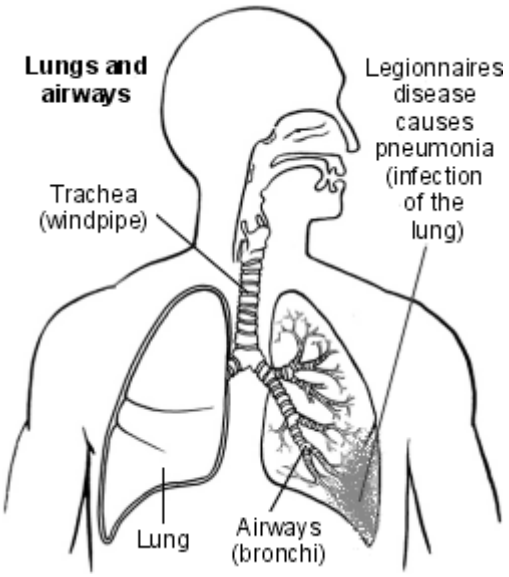

Figure.4. Human lung infected by legionnaire's disease.
http://www.patient.co.uk/showdoc/27000191/

1.7.2. Pontiac fever

Pontiac Fever, in contrast to legionnaire's disease, often goes undiagnosed and is generally a mild and a self-limiting influenza like upper respiratory tract infection. Headache, sore throat and cough, develop over a period of 6-24 hours and may be accompanied with neck stiffness and photophobia. The recovery period is only a matter of a few days (Pattison, N. et al. 2006).

1.8. Pathogenesis and Immunity

The disease causing *Legionella* are facultative intracellular parasites. Once inside the human lungs they can multiply in alveolar macrophages similar to their way of multiplying in nature in free-living amoebae (Fields, Benson et al. 2002). Macrophages are a type of white blood cells that engulf foreign materials. Macrophages play a key role in the immune response to foreign

invaders such as infectious microorganisms. Blood monocytes migrate into the tissues of the body and there mature into macrophages. Tissue macrophages are large irregularly shaped cells characterized by an extensive cytoplasm with numorous vaculoles. They have granular inclusions called lysosomes. Macrophages are phagocytes, acting in both innate immunity as well as adaptive immunity. Their main role is phagocytosis, that means engulf, and disposing of dead cells and cell debris as well as invading microorganisms, and to activate cell mediated immunity or adaptive immunity. When macrophages engulf microorganisms, the microorganisms become trapped in a vacuole, which then fuses with a lysosome. Within the lysosome, which contains bactericidal substances such as hydrogen peroxide, lysozyme, proteases, phosphatases, nucleases, and lipases digest the invader. Figure 5 shows how macrophage engulfs microorganisms and destroys them.

However, some bacteria, such as *Mycobacterium tuberculosis*, and *L. pneumophila* have become resistant to this method of digestion (Parham 2005). *Legionella* have membrane-bound proteins on the surface, receptors on the macrophages recognize and bind to these proteins in the invading bacteria and engulf them. The bacterium, when bound to the macrophage, penetrates the cell through endocytosis. The host cell is unable to kill the cell as the bacteria are able to inhibit phagosome-lysosome fusion, thus avoiding exposure to toxic superoxide, hydrogen peroxide, and hydroxyl radicals.

The organism then proliferates in the intracellular vacuole and produces proteolytic enzymes, phosphatase, lipase, and nuclease which all eventually lead to the death of the cell when the vacuole is lysed (Fields, Benson et al. 2002).

The macrophage infectivity potentiator protein gene *(mip*) has been identified as necessary for the multiplication of the bacteria within the alveolar macrophages. The *Mip* surface protein is a prokaryotic homolog of the FK506-binding proteins (Cianciotto and Fields 1992).

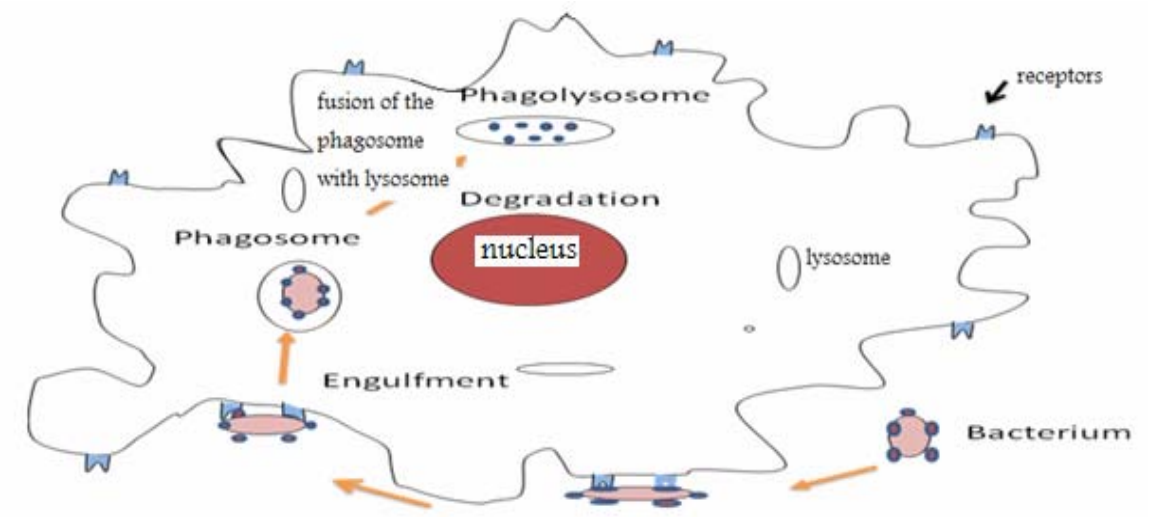

Figure.5. Macrophages respond to pathogens(figure modified from the original one)(Parham 2005).

1.9. Identification of *Legionella*

Rapid identification of the causative agent of disease, in this case *L. pneumophila*, is imperative as a delay in starting of proper treatment could result in increased mortality (Stolhaug and Bergh 2006).

1.9.1. Culture methods

In order to identify the disease causing agent it must be isolated in the laboratory. The standard procedure for isolating *Legionella* bacteria from the environment or from clinical samples is detailed in the (Brenner, Krieg et al. 2005) Figure 6. Shows the colonies of the *L. pneumophila* on the buffered charcoal yeast extract agar.

Legionella can be cultured on Buffered Charcoal Yeast Extract agar containing L-cysteine and iron (III). The cells form colonies on the agar surface after at least 2 days. The colonies are often white, purple to blue or lime green in colour but may be pink or red. Under ultraviolet light, colonies of several species (*L. bozemanii, L. gormanii, L. dumoffii, L. anisa, L. cherrii, L. steigerwaltii, L. gratiana, L. tucsonensis and L. parisiensis*) autofluoresce brilliant white; *L. rubrilucens* and *L. erythra* appear red. Colonies of *L. pneumophila* appear dull green often tinged with yellow. The colour of the fluorescence can help to differentiate colonies in samples containing different species of *Legionella*. The colonies have a characteristic ground-glass appearance when viewed with a low power stereomicroscope (Brenner, Krieg et al. 2005).

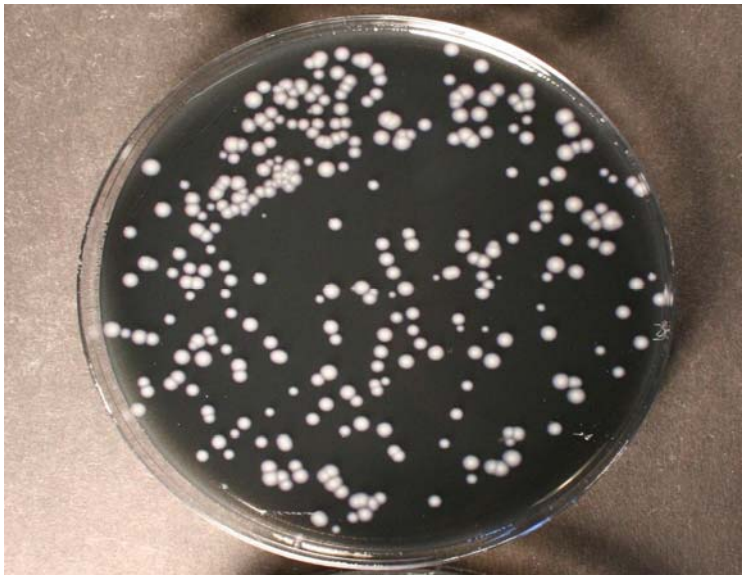

Figure.6. The colonies of *L. pneumophila* on the buffered charcoal yeast extract agar.

1.9.2. Other identification methods

Isolation and positive identification of *Legionella* in pure culture is considered as "gold standard" for detection of *Legionella*. However, due to its slow growth and fastidious nature other methods are becoming popular.

Several PCR analysis methods are present to target *Legionella* spp. and *L. pneumophila* genes ; These targeting, for example, the 16S rRNA gene, the 5S rRNA gene, the 23S – 5S spacer region, and the macrophage infective potentiator (*mip*) gene (Stolhaug and Bergh 2006). (Grimont, Grimont *et al.* 1985) developed the first nucleic acid probe for all species of the genus *Legionella* and for the species of *L.pneumophila.*

The direct fluorescent antibody technique has also been widely applied to the detection of *Legionella pneumophila* in clinical and environmental samples (Wilkinson and Brake 1982; Brenner, Krieg et al. 2005).

Legionella pneumophila was first detected in yolk sacs of embryonated hens`eggs infected with suspensions of tissues of guinea pigs by the staining method of Gimenez (Brenner, Krieg et al. 2005). A more selective staining method, i.e immunofluorescent staining is an extremely rapid method to discover *Legionella* from the respiratory tract and lung tissue specimens. Direct immunofluorescent test (DFT), using fluorescein isothiocyanate coupled to rabbit immunoglobulin to detect *Legionella* in clinical samples was developed by(Cherry, Pittman et al. 1978). An indirect immunofluorescent test is an alternative and more economical technique. An advantage of immunofluorescence tests is the possibility to detect non- culturable *Legionella* (Hussong, Colwell *et al.* 1987).

The sputum and urine are the first clinical samples used to detect *Legionella* antigens using enzyme-linked immunosorbent assay (ELISA) (Berdal, Farshy et al. 1979; Tilton 1979).

1.10. Typing of *Legionella*

When incidents of Legionnaire`s disease are identified epidemiologic and environmental investigation should start immediately to determine the source of transmission and to stop the spread. There is a need not only to identify the causative agent but to trace and prevent further distribution. Currently available methods suffer from a number of weaknesses such as their poor

reproducibility within and between laboratories and their inability to quantify the genetic relationships between isolates. Due to the public health and economic consequences associated with a false attribution of source of infection, it is vital that a reliable typing method exists.

1.10.1. Typing methods

Multilocus sequence typing (MLST) has been applied to *Legionella* and is currently promoted for widespread use by the European Working Group on *Legionella* Infections (EWGLI) (Gaia, Fry et al. 2005). This technique is based on the sequencing of internal fragments of selected housekeeping genes.

1.10.2. Multi Locus Variable Number of Tandem Repeat Assay (MLVA)

Recently, repeat based genotyping analysis for typing, identification and tracking purposes has become more common. Especially after bioterrorist attacks for example, anthrax attacks, in the USA there has been an increased interest in developing rapid and precise molecular identification and typing techniques (Lindstedt 2005).

In the past 5 years, multi-locus variable number tandem repeats analysis (MLVA) has been recognized as a modern, appropriate and flexible technique for molecular typing of several pathogenic bacteria such as *Bacillus anthracis (Keim, Price et al. 2000), Yersinia pestis (Adair, Worsham et al. 2000; Le Fleche, Hauck et al. 2001), Mycobacterium tuberculosis (Le*

Fleche, Fabre et al. 2002), Haemophilus influenza (van Belkum, Melchers et al. 1997), Francisella tularensis (Farlow, Smith et al. 2001), Xylella fastidiosa (Coletta-Filho, Takita et al. 2001), Staphylococcus aureus (Sabat, Krzyszton-Russjan et al. 2003), Salmonella enteric (Lindstedt, Heir et al. 2003; Liu, Lee et al. 2003), Escherichia coli 0157 and *Neisseria meningitidis* (Yazdankhah, Lindstedt et al. 2005; Liao, Li et al. 2006).

Minisatellite repeats were first discovered in 1980 by A.R. Wyman and R. White in humans (Wyman and White 1980). The first MLVA assay was used for human genotyping. In humans polymorphic properties of variable number tandem repeats (VNTRs) make them useful in areas such as paternity determination, origin testing and forensic medicine. Recently, minisatellite variability has been extensively studied in eukaryotes (Lindstedt 2005)). Studies are revealing that variations within microsatellites are associated with a number of human diseases for example- trinucleotide repeat disorders.

More recently it has been shown that the polymorphic properties of VNTRs in bacteria can be also used for high resolution typing (O'Dushlaine, Edwards et al. 2005). These regions carry significant information on evolution and genetic relationships of bacteria (O'Dushlaine, Edwards et al. 2005).

What are Tandem repeats?

In the last few years, a number of bacterial genomes have been fully sequenced and it has been discovered that many repetitive sequences exist in bacterial genomes, for example direct repeats, dyad repeats and inverted

repeats (Lindstedt 2005; Chang, Chang et al. 2007). Minisatelites and microsatelites are examples of direct tandem repeats (TR). These repeats can be gathered in one place or spread throughout the entire genome. Microsatellites are up to 6-bp in length while minisatellites are usually more than 5-bp in length. There is no significant difference between micro and minisatellites other than repeat unit size. Minisatellites are generally GC-rich tandem repeats. They are polymorphic due to variations in repeat copy number and repeat sequences also may vary (Bois 2003). These repeats are called variable number tandem repeats (VNTRS). The variability may, as in eukaryotes, be attributable to polymerase slipped-strand mispairing. At present it is unclear if these events are coincidental or dependent upon natural selection, yet there is no study that has fully explored the function of the VNTRs in prokaryotes.

Why does repetitive DNA vary?

There are many reasons for polymorphism in tandem repeats. Slipped strand mispairing is the best available scientific explanation for the variability of short sequence repeats. Since the number of repeat units varies from individual to individual, they have been used as DNA markers for molecular typing of several pathogenic bacteria such as *Bacillus anthracis, Yersinia pestis, Mycobacterium tuberculosis, Haemophilus influenza, , Staphylococcus aureus, Salmonella enteric, Escherichia coli 0157* and *Neisseria meningitidis*.

reference (Chang, Chang et al. 2007). It has been documented in many studies that the repeat number in VNTR is a strain- defining parameter (Par, Kantor et al. 1991).

What is replication slippage?

VNTRs are unstable units which undergo frequent variation in the number of units through slipped strand mispairing during DNA synthesis or for example double strand break repair, -replication slippage occasionally generating a new length variant. In replication slippage, stretches of fairly short arrays of repeat units, when being copied by the DNA polymerase, may engage in illegitimate base pairing. This forces the polymerase to introduce or delete individual repeat units.The result is that the new polynucleotide has more or fewer, respectively, of the repeat units Figure7. Shows how replication slippage occurs in a two unit repeat minisatellite. Slippage has occurred during replication of the original strand, inserting an additional repeat unit into one of the newly synthesized strand. Later this new strand replicates and gives a one unit longer minisatellite than that of the initial strand. This is the main reason why microsatellite sequences are so variable. Undesirable replication can also result in small numbers of extra nucleotides being inserted or some nucleotides in the template not being copied (Brown 2007).

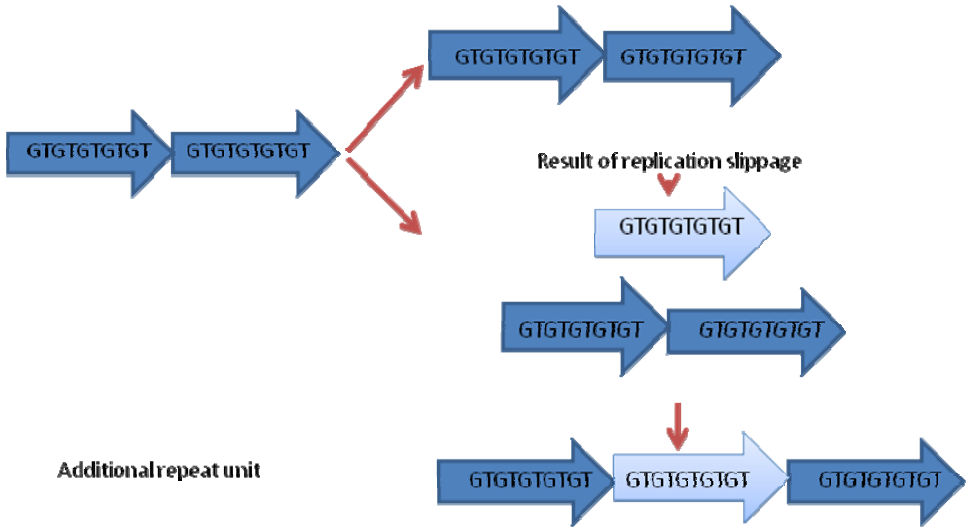

Figure 7. How replication slippage occur in a two unit repeat minisatellite.

During DNA replication of a circular bacterial genome, recombination can occur between two daughter double helices that have simultaneously been synthesized by DNA replication. Figure.8 shows how recombination occurs during replication of a bacterial genome (Brown 2007).

During replication of a bactyerial genome, some times an extra repeat unit is inserted in the newlysynthesized strand by recombination.(figure 8.)

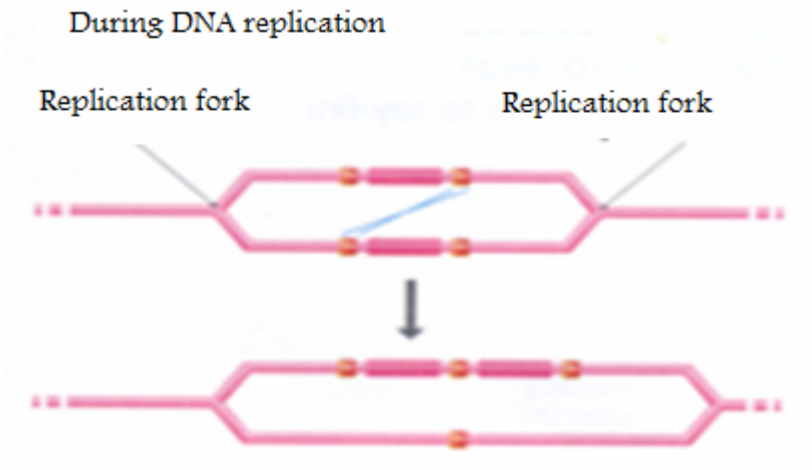

Figure 8. Recombination during replication of a bacterial genome and creation of an extra repeat unit in the newly synthesized strand (Brown 2007).

What is double strand break repair?

DNA is a delicate thread that often breaks when exposed to free radicals generated during essential metabolic process or to natural mutagens. When it does, the cell must find a way to bring back broken ends together in order to continue its cycle of replication. Cells possess an array of ways to rejoin broken DNA ends. The two broken ends must be protected from further degradation, if the broken region is not protected which could result in a deletion appearing at the repaired break point (Brown 2007).

In some microorganisms these repeat differences influence genome function. In some bacteria and yeast TRs are present within the protein sequence, some studies are showed that changes in repeat numbers can alter the amino acid sequence of the corresponding protein, and for example, changes in cell wall proteins due to changes in the number of TRs can cause alterations in immunogenicity, adhesion and pathogenesis (Jordan, Snyder et al. 2003; Sylvestre, Couture-Tosi et al. 2003).

Analysis of tandem repeats

The multiple locus variable number of tandem repeats assays are based on the detection of tandem repeats sequences in the bacterial genome.
This assay consists of four steps.

1. DNA extraction

2. PCR amplification of specific loci using specific sets of primers spanning the repeat region.

3. Separation and detection of PCR products.

4. Determining the sample genotype.

After PCR amplification another step is required to analyse the PCR product length. There are many analysis methods available to determine the PCR product length such as agarose gel electrophoresis, Lab-on-a chip systems and advanced mass spectrometry. Some of these methods are costly, some of them are time consuming and some of them are not very accurate.

Initially, agarose gel electrophoresis was the most widespread method to separate and determined the size of amplified PCR products. In agarose gel electrophoresis size determination of the fragments run on a gel is done by comparison with a size marker, using a Software program for example, Gene Tools, to determine the size of the fragments. In Gene Tools a cursor (a moving marker on a computer screen that marks the point at which band will be appear) is placed on a specific band and the band size is estimated by comparison with a size marker. When this cursor is placed in the middle of a fat band resulting from DNA overloading, the observed size will not be accurate. The nature of the gel matrix can also influence the fragment migration and may lead to incorrect size determination.

Capillary electrophoresis (CE) is one way to determine PCR product lengths. Capillary electrophoresis is based on the principles of electrophoresis. Electrophoresis means the differential movement of ions by attraction in an

electric field. A capillary is a narrow-bore tube, normally from 25 to 100µm inner diameter. The speed of movement or migration of solutions in CE is determined by their size and charge. The detector response is usually UV-visible absorbance or fluorescence. The data output from CE is called an electropherogram, which is similar to a chromatogram. The instrument used to carry out electrophoresis consists of an autosampler, a detection module, a high-voltage power supply, the capillary and, a computer (Frazier, Ames et al. 2000).

The benefits of CE include the ability to analyze multiple PCR products in the same capillary. Because of its high capacity with the use of 96 well plates 96 samples to be analyzed simultaneously. Separation and size calling are automated, allowing time saving and reduce the costs. Additionally as the fluorescence signal is laser detected it makes this method highly accurate and reproducible (Smith and Nelson 2004). This method is fast, less labour-intensive than agarose gel electrophoresis and sensitive. For these reason it is developing fast and playing an increasingly important role in the typing, identification and tracking purposes.

Pourcel *et al*. (2003; 2007) developed a MLVA method to address the genetic diversity of *L. pneumophila* based on eight minisatellite markers. The PCR product lengths were resolved using agarose gel electrophoresis. In the present study we employed capillary electrophoresis to separate and determine the amplified PCR product lengths for the MLVA-8 markers. As far as we are aware this is the first time that capillary electrophoresis has been used as part of the MLVA analysis of *Legionella pneumophila*.

1.11. Aim of this study

The aim of this study was to optimise a high resolution genetic typing method known as MLVA for use on capillary electrophoresis and to apply this method to a number of environmentally and clinically isolated *Legionella pneumophila* bacteria. It is hoped that this technique shall be used during future studies which aim to track *Legionella pneumophila* during outbreak situations.

2. Materials and Methods

2.1. Materials

All pre made chemicals materials, kits, solution, buffers, used during lab work, and suppliers information, are specified in the table 2 . Salts, simple organic compounds and other routinely used chemicals were of pro analysis quality and suppliers are not listed. Filter pipette tips and gloves were used all the time (for quantifying DNA and gel loading pipette tips without filters were used). The water used was MilliQ ultrapure water. All the lab work was carried out in under sterile conditions using sterile materials. Ultrapure water was used as a negative control.

Table 2. List of premade chemicals and suppliers

Method	chemicals	Supplier (manufacture)
PCR	PCR Buffer	Eurogen tec
PCR	dNTP mix	Saveen&Werner
PCR and sequencing	Primers	ABI and MWG
PCR	Hot Gold Star Taq	Eurogen tec
PCR	MgCl2	Eurogen tec
CE	Genescan-1200 LIZ Standard	Applied Biosystems
CE	Formamide	Applied Biosystems
AE	Ladder	Fermentas
AE	Gel loading buffer	Fermentas

PCR- Polymerase chain reaction. CE- Capillary Electrophoresis. AE-Agarose gel Electrophoresis

2.2. Methods

Bacterial strains used in this study are listed in the table 3 and 4

Legionella pneumophila strains

The bacterial strains used during optimization experiments and their origins are listed in Table.3. The strains represented in this table were originally used in (Pourcel, Visca et al. 2007) paper as part of their proficiency panel.

Table 3. *Legionella pneumophila* strains used for optimization of MLVA method (Table modified from (Pourcel, Visca et al. 2007)

Culture no.	Source	Original designation	serogroup	Country of origin	Epidemiological relatedness
NCTC11199	Health Protection Agency UK	Philadelphia-1	1	United States	Unrelated
EUL 146	EWGLI	Paris	1	France	Unrelated
EUL 160	EWGLI	Lens	1	France	Unrelated
EUL 025	EWGLI	L3	1	France	Unrelated
EUL 048	EWGLI	006/96	1	Spain	Clinical isolate from same patient as 17/96
EUL 056	EWGLI	17/96	1	Spain	Clinical isolate from same patient as 006/96
EUL 121	EWGLI	R4-Augsburg1	1	Germany	Unrelated
EUL 137	EWGLI	Corby CA	1	Uk	Unrelated
EUL 153	EWGLI	LC 0202	6	Uk	Unrelated
EUL 154	EWGLI	LC 0569	8	Uk	Clinical isolate related to LC0606
EUL 155	EWGLI	LC 0606	8	Uk	Environmental isolate related to LC0569
EUL 156	EWGLI	LC 0348	10	Belgium	Unrelated
EUL 157	EWGLI	H04 280 0510	1	Uk	Unrelated

NCTC, National Collection of Type Cultures, UK;EUL, European Union Legionella

Table 4. Bacterial strains obtained from Norwegian Institute of Public Health

UIO code	Serogroup	year	Legionella species	Source
000	1		*L. pneumophila*	Clinical isolate
001	-		*L. bozemanii*	Clinical isolate
002	-		*L. Londiniencis*	Clinical isolate
003	-		*L. pneumophila,*	Clinical isolate
004	-		*L. bozemanii*	Clinical isolate
005	-		*L. micdadei*	Clinical isolate
006	-		*L. pneumophila*	Clinical isolate
007	-		*L. pneumophila*	Clinical isolate
008	-		*L. pneumophila*	Clinical isolate
009	-		*L. anisa*	Clinical isolate
024	-	2007	*L. erythra*	Control strain
025	-	2006	*L. oakridgensis*	Biodam
026	1	2005	*L. pneumophila*	River Glomma
027	-	2005	*L. anisa*	Environmental swab
028	2 - 14	2007	*L. pneumophila*	Snow melt water
029	-	2006	*L. pneumophila*	Biodam
030	-	2006	*L. londiniencis*	Biodam
031	1	2005	*L. pneumophila*	Cooling tower
032	-	2005	*L. anisa*	Environmental swab
033	1	2005	*L. pneumophila*	Clinical isolate
034	1	2007	*L. pneumophila*	Snow melt water
035	1	2007	*L. pneumophila*	Snow melt water
036	-	2006	*L. pneumophila*	Biodam
037	1	2005	*L. pneumophila*	Clinical isolate

2.2.1. DNA Isolation

DNA of all the strains analysed in this study were Isolated at Norwegian Institute of Public Health.

2.2.2 Quantification of DNA

Isolated DNA was quantified using spectrophotometry (Nano Drop, on model –ND-1000 spectro photometer). DNA concentration was calculated based on absorption at 260nm. DNA was diluted with water to a final concentration of 10ng/µl.

2.2.3. Polymerase Chain Reaction (PCR)

The purpose of a PCR-Polymerase Chain Reaction is to dramatically increase the number of copies of known fragment of DNA. In the non limiting phase of the reaction the increase in the number of copies of the fragment is exponential. There are three main steps in a PCR reaction which are repeated for many cycles. Step one, denaturation, during this step the double strand melts open to single stranded DNA. Step two annealing, in this step where primers anneal to the template DNA. Finally step three extension where the DNA polymerase attaches and copies the DNA template. All this process is carried out on an automated cycler. Verification of the PCR reaction can be done by analyzing the PCR product on an agarose gel. Presence or absence of the PCR amplification, and its size in base pairs (bp) can be verified in this manner (Brown 2007).

2.3. Optimization of the PCR for MLVA analysis

Initially the PCR conditions used by Pourcel *et al.* (2007) for the same bacterial strains and primers were used. However, our results showed that using these conditions were not satisfactory; therefore the PCR was optimized further (detailed below). PCR primers used in the experiments are listed in Table.6. The DNA was amplified individually for each primer set. Forward primers were labeled with a fluorescent tag.

2.3.1. Magnesium chloride ($MgCl_2$) concentrations

This was the first parameter that was optimized after obtaining failed or week weak PCR products. To test the influence of $MgCl_2$, PCR reactions with different $MgCl_2$ concentrations were performed with increasing $MgCl_2$ concentrations from 1.5 to 5mM.

2.3.2. DNA polymerase source

Two DNA polymerase enzymes were tested for their efficiency during PCR optimization. Reactions were carried out using GoTaq (Promega) and Hot Gold Star (Invitrogen).

2.3.3. PCR reaction volume

Total PCR reaction volumes between 10 and 25µls were applied.

2.3.4. Optimized PCR reaction mixture

PCRs were performed in 25µl PCR reaction mixtures according to Table.5.

Table 5. Optimal concentrations of the compounds for PCR reactions using primers for MLVA analysis

No	Solution for PCR	Stock	Final concentration	Amount per reaction (µl)
1	ddH2O	-	-	13.375
2	PCR Buffer	10x	1x	2.5
3	MgCl2	25mM	3.5mM	3.5
4	dNTP	2mM	0,2mM	2.5
5	Forward primer	10µM	0,4 µM	1
6	Reverse primer	10 µM	0.4 µM	1
7	Hot Gold Star Taq DNA polymerase	5U/ µl	0.025U\ µl	0.125
8	DNA Template	-	-	1
	Master mix per reaction	-	-	24
	Total per reaction	-	-	24

ddH2O- double distilled water,

Table.6. PCR primers used for MLVA analysis of *Legionella pneumophila* bacteria

Primer Set	Fluorescent Label	Colour	Wavelengths	Sequence	Allele size range(bp)
Lpms1_bF	VIC	Green	EX-538. EM-552	ACGAGCATATGACAAAGCCTTG	475-633
Lpms1_bR				CGGATCATCAGGTATTAATCGC	
Lpms3F	FAM	Blue	EX-494. EM-518	CAACCAATGAAGCAAAAGCA	845-941
Lpms3R				AGGGGTTGATGGTCTCAATG	
Lpms13F	FAM	Blue	EX-494. EM-518	CAATAGCATCGGACTGAGCA	236-548
Lpms13R				TGCCTGTGTATCTGGAAAAGC	
Lpms17F	PET	Red	_	CAGCTCACCCCGTATCACTT	259-278
Lpms17R				TAACATCAATGACCGCGAAA	
Lpms19_bF	NED	Yellow	EX-546. EM-575	GAACTATCAGAAGGAGGCGAT	173-194
Lpms19_bR				GGAGTTTGACTCGGCTCAGG	
Lpms33F	NED	Yellow	EX-546. EM-575	ACCACAGCAGTTTGAACATAAT	227-727
Lpms33R				GGGAGAAGTTATAGATCTATTCG	
Lpms34F	VIC	Green	EX-538. EM-552	GAAAAGGAATAAGGCGCAGCAC	209-459
Lpms34R				AAACCTCGTTGGCCCCTCGCTT	
Lpms35F	PET	Red	_	CTGAAACAGTTGAGGATGTGA	
Lpms35R				TTATCAACCTCATCATCCCTG	202-724

F on the end of the primer name denotes forward primer and R denotes reverse. Only forward primers were fluorescently labeled. EX-Excitation wavelength. EM- Emission wavelength.

2.3.5. PCR cycling conditions

Annealing temperatures and PCR cycling conditions (which included cycle number) were optimized for the primer sets used. Touchdown PCR (TD) was used for some of the primer sets to reduce unspecific amplification. In Table.7. lists the optimized PCR programs used for the various MLVA primer sets.

Touchdown PCR- Touchdown PCR is a method which reduces unspecific amplification. To eliminate the unspecific amplification an annealing temperature that is higher than the optimum temperature in early PCR cycles is used. The annealing temperature is decreased by 0.5 °C every cycle until a specified annealing temperature is reached. This temperature is then used for the remaining number of cycles. This allows eliminating unspecific amplification products. (http://www.bio.net/bionet/mm/methods/1992-July/000055.html).

Table .7. Optimized PCR programs used for the various MLVA primer sets

Primer sets	Initial denaturation	Denatu-ration	Annealing	Elongation	Final elongation	Number of cycles
Lpms 13, 34, 1b, 3 and 35	95°C 10 min	95°C 15 sec	TD 75-65°C[1] 30 sec	72°C 45 sec	72°C 7 min	35
Lpms 19 and 33	95°C 10 min	95°C 15 sec	TD 70-60°C[2] 30 sec	72°C 45 sec	72°C 7 min	35
Lpms 17	95°C 10 min95°C 15 sec	95°C 15 sec	46.0°C	72°C 45 sec	72°C 7 min	40
Lpms 17	95°C 10 min95°C 15 sec	95°C 15 sec	57.8°C	72°C 45 sec	72°C 7 min	
Mip	95°C 10 min	95°C 15 sec72°C 7 min		72°C 45 sec	72°C 7 min	40

TD-touchdown PCR where annealing temperature is decreased by 0.5°C very cycle until 20 cycle.

[1]After 20 cycles annealing temperature remained constant at 65^{o}C for the remainder of the cycles.

[2]After 20 cycles annealing temperature remained constant at 60^{o}C for the remainder of the cycles

2.4. Agarose gel electrophoresis

Initially agarose gel electrophoresis was used to verify the results of PCR reactions. After PCR, the PCR products were examined by electrophoresis on a 1.5 % agarose gel. 10 µl of ethidium bromide was added to the liquid agarose gel mixture before it was cast. A gel 'comb' was used to make wells within the gel during the casting period. The gel was allowed to set for 10 minutes. 10 µl of PCR product was mixed with 2 µl of gel loading buffer and loaded onto the gel. The gel was run for 50 minutes at 90V. After electrophoresis, the agarose gel was placed under UV light and a polar camera was used to take a picture of the DNA separation pattern. To determine the size of the PCR fragments was done by comparison with a commercial size marker (Fermentas). To demine the accurate size a software program gene tools, was used. The band size was determined by the way the cursor was placed on a band and band sizes were estimated (Brown 2007).

2.5. MLVA analysis by capillary electrophoresis (CE)

Analysis of PCR products can also be carried out on capillaries as capillary electrophoresis. Capillary electrophoresis is based on the principles of electrophoresis. A Capillary is a narrow-bore tube, normally from 25 to 100µm inner diameter. The speed of movement or migration of solutions in CE is determined by their size and charge. The detector response is usually UV-visible absorbance or fluorescence. The data output from CE is called an electropherogram, which is similar to a chromatogram. The instrument used to carry out electrophoresis (Applied biosystems 3730 DNA analyzer in this case)

consists of an autosampler, a detection module, a high-voltage power supply, the capillary and, a computer to control everything (Blessum, Jeppsson et al. 1999).

2.5.1. Optimization of capillary electrophoresis

Optimization of the capillary electrophoresis was performed to be able to establish a method, which could be reproducible, reliable and give certain results through measurement of absorption and visualization of the fragments.

2.5.2. Optimization of internal size standard

Liz1200 size standard (Applied Biosystems) was used as an internal standard for capillary electrophoresis. This size standard allowed size determination up to 1200bp. Since it is currently not commercially available, we were selected as early users of this product by Applied Biosystems. Different amounts of the size standard were run to test the most suitable size range and Liz1200 size standard was used for sizing fragments between 20 and 1200 base pairs.

2.5.3. Optimization of the amount of PCR product loaded on the Capillaries for electrophoresis.

Different dilutions of the PCR products were run on the CE to test for the most appropriate concentration of DNA to be loaded. Four dilutions of the PCR products were tested, 10x, 50x, 100x, 20x.

2.5.4. Optimization of the running parameters of the DNA analyzer

Different 'Run times' (time the machine takes to detect samples from the start of the electrophoresis) were tested. Injection times were also varied. Optimal settings were run voltage-8V

Runtime-6200sec.

2.5.6. Optimized capillary electrophoresis method

For running on the DNA analyzer PCR products were pooled in two separate mixtures. This resulted in the multiplex analysis of the PCR products of each mixture in one capillary. The PCR products within one reaction mixture had been amplified with different fluorescent labels to allow separate detection of each product on the capillary. The first sample (Panel I) contained 1µl of PCR products for markers Lpms 13, 19b, and 1µl for each of the two Lpms 17 PCR`s. This sample was diluted to 100µls. In the second sample (Panel II) 1µl of PCR`s for markers Lpms1b, 33, and 35 and 3µl for marker Lpms 3 were mixed

and also diluted to 100µl. DNA samples were prepared for the capillary electrophoresis by adding 1µl of a diluted PCR product mix to 8.8 µl of formamide and 0.2 µl of the Liz1200 size standard (used for sizing fragments between 20 and 1200 base pairs) in the 96 well optical plates. The samples were heat denatured at 95 °C for 3 min and then cooled on ice. Fragment analysis was performed on the 3730xl sequencer (Applied Biosystems) Using POP-7 polimer.The samples were then injected at 1.6 volts for 10 s and separated at 8 K volts for 6200sec with a run temperature of 35 °C. The results were analyzed using the Genemapper software Version 3.7 (Applied Biosystems).

2.6. Sequencing of *mip* gene of *Legionella* spp.

The sequencing method used was based on the Sanger- sequencing method. This method employs the dideoxynucleotiedes chain termination that means the use of dideoxynucleotiedes in addition to the normal nucleotides present in DNA. Actually dideoxynucleotiedes contain a hydrogen group on the 3`carbon instead of a hydroxyl group. These modified nucleotides, when integrated into a sequence, prevent the addition of further nucleotides or stop further elongation. This occurs because a phosphodiester bond cannot form between the dideoxynucleotiede and the DNA chain elongation is terminated (Brown 2007).

All the bacterial strains were amplified using MIP gene specific primers listed in the Table 8 using the PCR program for Mip listed in the Table 7 Unpurified PCR products were used for the sequence reaction which was set up as follows:

- 2 µl PCR product (unclean), 1 µl of the forwared/reverse primer (in two different reactions) and dH2O to a total reaction volume of 10 µl.
- Sequencing was performed at the on the ABI sequencer (3730xl) by the sequencing lab at the Biology institute UIO.
- DNA sequences viewed using Contig Express (a component of vector NTI Advance 10.3.0-Invitrogen Corp) sequence analysis software.

Table .8. Primers used for MIP gene analysis

Primer	Type	Sequence
Legmip_forward_M13Rev	PCR	CAGGAAACAGCTATGACCGGGRATTVTTTATGAAGATGARAYTGG
Legmip_revers :M13(-21)	PCR	TGTAAAACGACGGCCAGTTCRTTNGGDCCDATNGGNCCDCC
Legmip_f_M13Rev	Sequencing	CAGGAAACAGCTATGACC
Legmip_r:M13	Sequencing	TGTAAAACGACGGCCAGT

Analysis was carried out using the BLAST program at the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST). The Mip gene sequences were blasted against the protein database using a translated nucleotide query.

3. Results

The MLVA method for genotyping of *Legionella pneumophila* was developed for capillary electrophoresis and tested within this study. Eight variable number tandem repeat (VNTR) markers were used. The markers were labelled as Lpms1b, Lpms3, Lpms13, Lpms17, Lpms 31, Lpms33, Lpms34, and Lpms35. Lpms stands for *L. pneumophila* minisatellite (Pourcel, Vidgop et al. 2003).Figure 9 shows the positions of the eight markers on the genome of *L. pnemophila* strain Philadelphia.

The bacterial strains used during the experiments were part of the proficiency panel including reference strains -Philadelphia, Paris, Lens (originally used in the Pourcel *et al.* 2007 paper) and 24 strains obtained from the Norwegian Institute of Public Health.

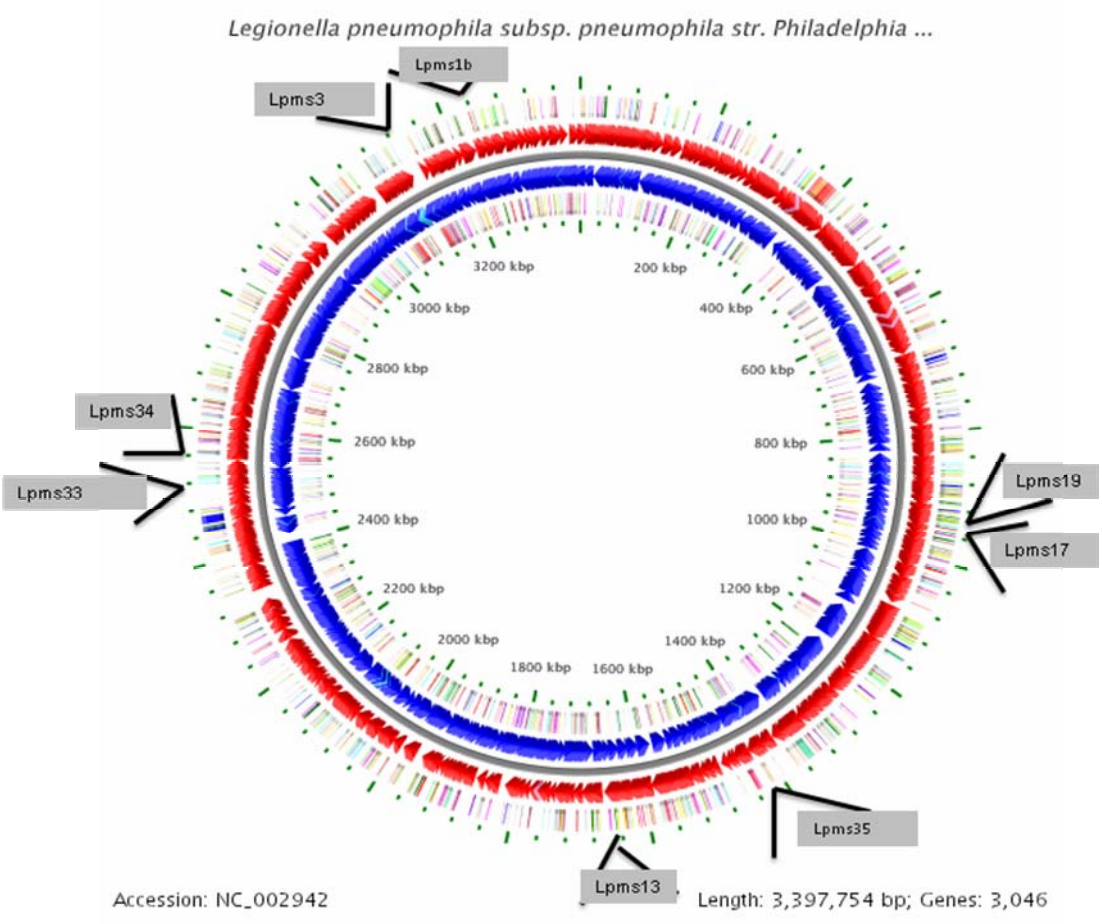

Figure 9. Position of the eight markers. The figure is modified from the original one located at http://wishart.biology.ualberta.ca/BacMap/cgview_linked_maps/NC_002942/index.html

3.1.1. Optimization of PCR

To improve the efficiency of minisatellite analysis, reaction conditions needed to be optimized to avoid nonspecific amplification products. Nonspecific products can be detected as smears or faint bands in addition to the bands

that are of interest. Non-specific bands can have a negative effect on capillary electrophoresis analysis because they appear as extra peaks. PCR optimization was performed by modifying reaction conditions including the annealing temperature, concentration of $MgCl_2$, PCR reaction volume and the source of DNA polymerase.

3.1.2. Optimizing the $MgCl_2$ concentration

The importance of an optimum magnesium chloride concentration for PCR is well recognised. To test the effect of $MgCl_2$, PCR reactions were performed with increasing concentrations from 1.5 to 5mM. Figure 10 shows the effect of the $MgCl_2$ concentration during PCR amplification. The lowest $MgCl_2$ concentration (1.5mM) failed to yield visible bands, and the highest $MgCl_2$ concentration 5.00 mM gave additional unspecific amplification products. Taking into consideration the amount of PCR product, the optimum magnesium concentration was between 3.00 and 4.00mM. The $MgCl_2$ concentration chosen for subsequent reactions was 3.5mM.

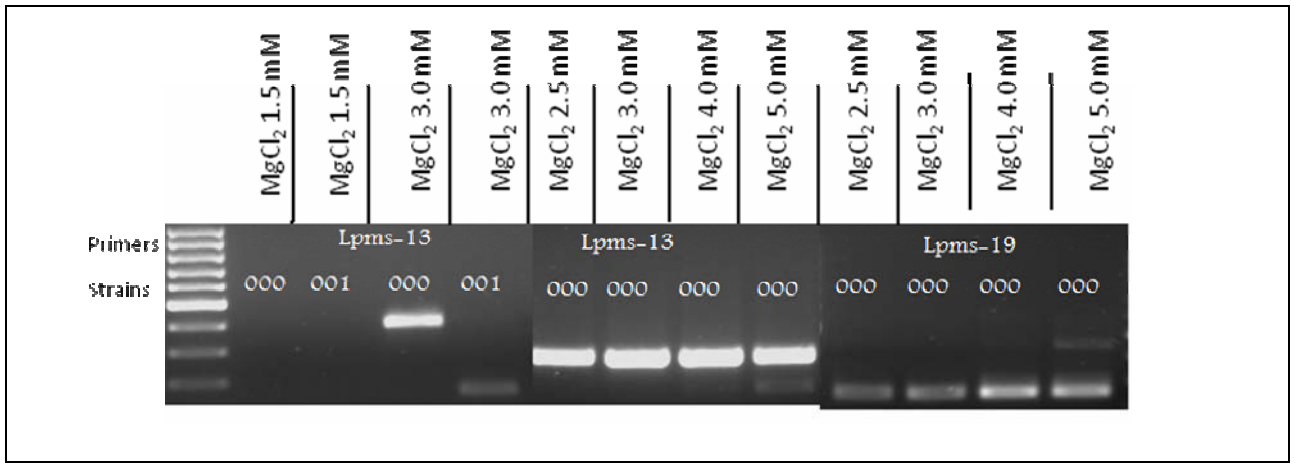

FIGURE 10. Effect of $MgCl_2$ concentrations
Gel electrophoresis of PCR products generated using different $MgCl_2$ concentrations for primers Lmps 13 & Lpms-19 on *L. pneumophilia* genomic DNA obtained from the Norwegian Institute of Public Health

3.1.3. Optimisation of DNA polymerase and PCR reaction volume

When Go Taq DNA polymerase was used in the experiments background bands were observed, as can be seen from figure 11A. Using Hot gold star polymerase the background bands were reduced.

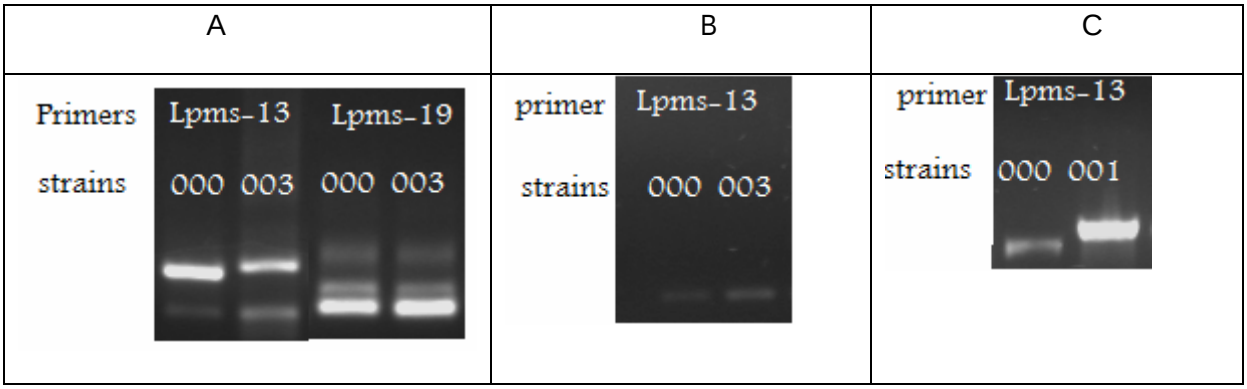

FIGURE 11. Effect of DNA polymerase and reaction volume

A. Gel electrophoresis of PCR products for primers Lmps 13 & Lpms 19 on strains 000 and 003 using Go Taq DNA polymerase (PCR reaction volume25 µl) **B.** Gel electrophoresis of PCR products for primers Lmps 13 on strains 000 and 003 using Hot Gold Star DNA polymerase (PCR reaction volume10 µl)**C**. Gel electrophoresis of PCR products for primers Lmps 13 on strains 000 and 001 using Hot Gold Star DNA polymerase (PCR reaction volume25 µl)

Go taq DNA polymerase is active at room temperature and this can be a reason for unspecific amplification. Another polymerase enzyme, Hot Gold Star (Eurogen tec) was tested. Hot Gold Star enzyme is not active at room temperature, having an activation temperature of 95^0C.

The Hot Gold Star DNA polymerase reduced the unspecific amplifications as shown in figure 11B. As only weak PCR products were obtained in the reaction the reaction volume was varied to test its effect. The PCR reaction volume influenced the outcome; 25µL PCR reaction volume gave significantly more yield than 10 µL reaction volume (figures 11.B and 11.C illustrate the influence of the reaction volume). Therefore it was decided to use 25 µl PCR reaction volume for subsequent reactions.

3.1.4. Touchdown PCR

Under normal PCR conditions, background smears were observed on the gel and many peaks in the capillary electrophoresis with one repeat size different in length (this is shown in page-64 Figure-19 (top). To reduce this effect Touchdown PCR was tried (figures 12 - 14). During touchdown PCR the annealing temperature is reduced by 0.5°C every cycle until a certain annealing temperature has been reached. The remaining cycles are all carried out at this initial temperature.

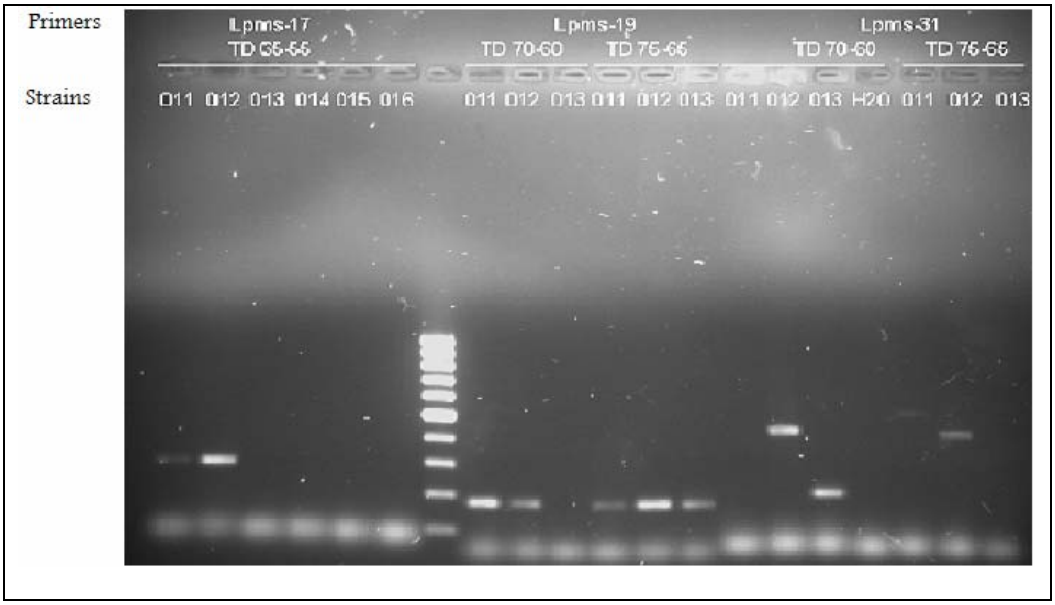

FIGURE 12. Effect of touch down PCR
Gel electrophoresis of Touch down PCR products for primers Lmps 17, 19 and 31 on reference strains analyzed on an agarose gel (90v for 60min)

TD-touchdown PCR where annealing temperature is decreased by 0.5°C every cycle for 20 cycles. TD-65-55 -After 20 cycles annealing temperature remained constant at 55°C for the remainder of the cycles. TD-70-60-After 20 cycles annealing temperature remained constant at 60°C for the remainder of the cycles. TD-75-65 -After 20 cycles annealing temperature remained constant at 65°C for the remainder of the cycles

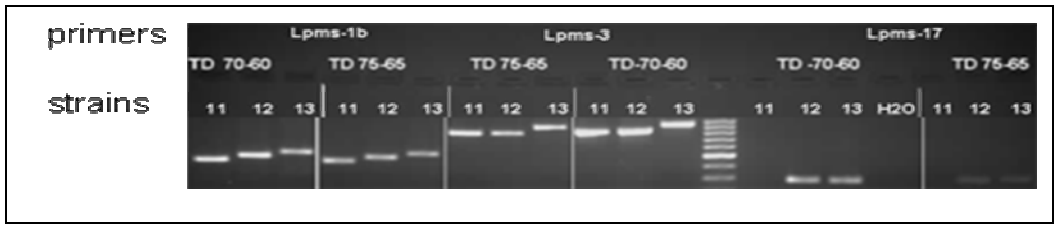

FIGURE-13: Effect of Touchdown PCR.
Gel electrophoresis of Touch down PCR products for primers Lmps 1b, 3 and 17 on reference strains analyzed on an agarose gel (90v for 60min)
TD-touchdown PCR where annealing temperature is decreased by 0.5°C every cycle for 20 cycles. TD-65-55 -After 20 cycles annealing temperature remained constant at 55°C for the remainder of the cycles. TD-70-60-After 20 cycles annealing temperature remained constant at 60°C for the remainder of the cycles. TD-75-65 -After 20 cycles annealing temperature remained constant at 65°C for the remainder of the cycles. In this figure strain 11, 12 and 13. refers to strain 011, 012 and 013 as detailed in table 4.

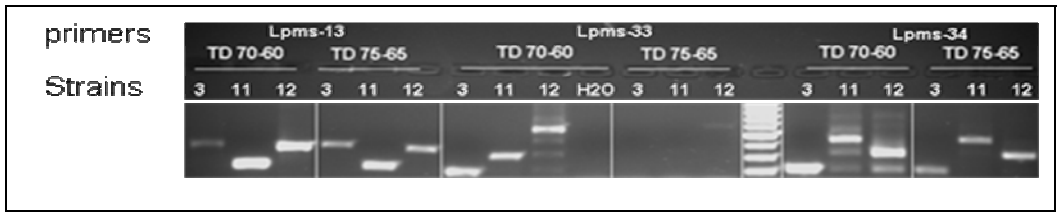

FIGURE 14. Effect of Touchdown PCR.
Gel electrophoresis of Touch down PCR products for primers Lmps 13, 33 and 34on reference strains analyzed on an agarose gel (90v for 60min)
TD-touchdown PCR where annealing temperature is decreased by 0.5°C every cycle for 20 cycles. TD-65-55 -After 20 cycles annealing temperature remained constant at 55°C for the remainder of the cycles. TD-70-60-After 20 cycles annealing temperature remained constant at 60°C for the remainder of the cycles. TD-75-65 -After 20 cycles annealing temperature remained constant at 65°C for the remainder of the cycles. In this figure strain 3,11and 12 refers to strain 003, 011 and 012 as detailed in table 4.

The touchdown PCR- with a temperature range of 75-65 °C in the first 20 cycles gave a good result for Lpms 13, 34, 1b, 3 and 19 primer sets, which is shown in figures 12, 13 and 15. For primer set 34 a faint smear and unspecific bands were still visible on the gel as shown in Figure.14. Touchdown PCR-temperatures using 70-60 °C (initial anneling temperature was 70 °C, later the annealing temperature was decreased by 0.5 °C every cycle. After 20 cycles

annealing temperature remained constant at 60 °C for the remainder of the cycles), improved the result for Lpms-33 although there still was a visible smear on the agarose gel. Touchdown PCR using 75-65°C (initial annealing temperature was 75 °C, later the annealing temperature was decreased by 0.5 °C every cycle. After 20 cycles the annealing temperature remained constant at 65 °C for the remainder of the cycles gave no bands as shown in figure14. The following conditions were chosen as optimal based on these experiments: touchdown PCR temperatures using 75-65 °C (initial start annealing temperature was 75 °C, later the annealing temperature was decreased by 0.5 °C every cycle. After 20 cycles the annealing temperature remained constant at 65 °C for the remainder of the cycles) for Lpms 13, 34, 1b, 3 and 19 primer sets, and temperatures using 70-60 °C (initial annealing temperature was high70 °C, later the annealing temperature is decreased by 0.5 °C every cycle. After 20 cycles the annealing temperature remained constant at 60 °C for remainder of the cycles) for Lpms-33, 31and 19.

3.1.5. Gradient PCR for the optimisation of Lpms primer set -17

Problems with primer set Lpms-17 were observed during tests with touchdown PCR(gave no amplification for some strains). To solve this (to get amplification for all strains) a number of annealing temperatures were tested during gradient PCR.

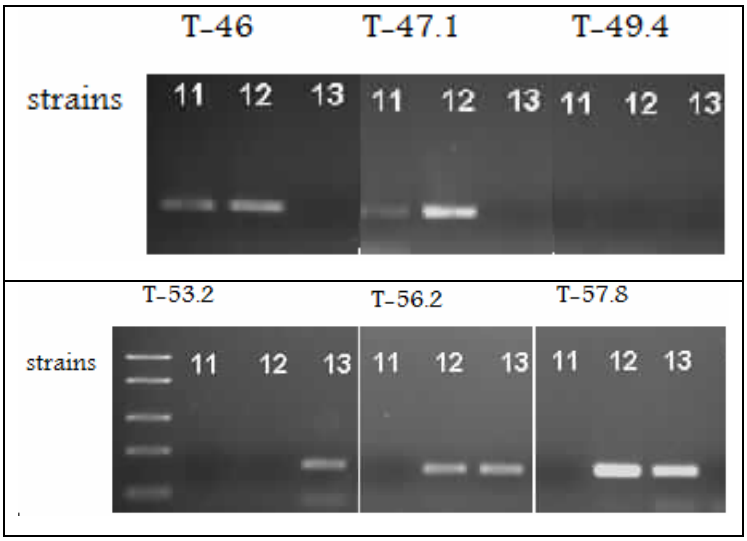

FIGURE 16. Gradient PCR to optimise Lpms primer set-17

Gel electrophoresis of PCR products for primer Lmps 17 on 3 strains using different annealing temperatures. T-temperature in °C. In this figure strain 11, 12 and 13. refers to strain 011, 012 and 013 as detailed in table 4.

For strain 013 bands were observed at 53.2°C, 56.2°C and 57.8°C as can be seen from figure 16. bottom. For strain 012 bands were observed at 46.0°C, 47.1°C, 56.2°C, 57.8°C and 58.0°C (figure 16). For strain 011 bands were observed at 46.0°C and 47.1°C (figure 16 (top)). These results imply that sequence variation is present within the strains of *L. pneumophila* where the primers anneal. In order to make sure amplification for this marker was obtained, it was decided to use both 46.0°C and 57.8°C for all strains. When the reactions were optimized and one band observed on the agarose gel, it usually resulted in only one band being detected in the capillary electrophoresis.

3.2.MLVA genotyping by agarose gel and capillary electrophoresis

It is important to make sure that allele assignments inferred from agarose gel electrophoresis are compatible with capillary electrophoresis. For this purpose the length of the eight markers of part of the proficiency panel including reference strains -Philadelphia, Paris, Lens (originally used in the Pourcel *et al.* 2007 paper) were first analyzed using agarose gel electrophoresis followed by capillary electrophoresis. The reproducibility of this method was examined by comparing expected values (from Pourcel *et al.* 2007) with fragment sizes obtained by both capillary and gel electrophoresis separation.

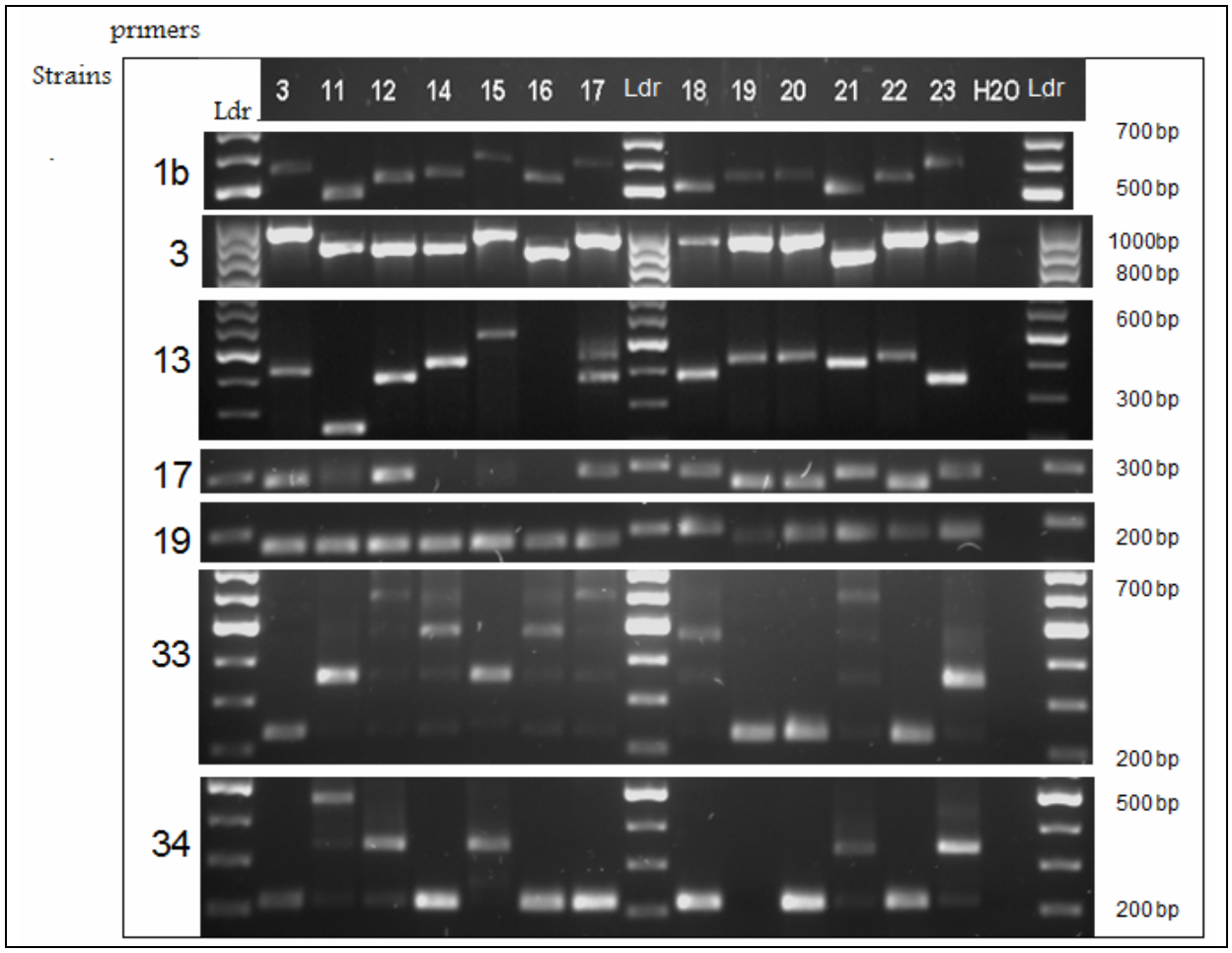

Figure 17. Aagrose gel electrophoresis
Gel electrophoresis of PCR products for primers Lmps 1b, 3, 13, 17, 19, 33, 34, on 13 strains analyzed on an agarose gel
Ldr- ladder
In this figure strain 3, 11, 12,14, 15, 16, and 17. refers to strain 003, 011, 012, 014, 015, 016 and 017 as detailed in table 4.

3.2.1. Agarose gel electrophoresis

The PCR reaction products for all eight Lpms set on 13 different strains of *L. pneumophila* were run on agarose gels and their size determined using gel image analysis software, the results are presented in figure 17.(in a few cases smears and unspecific amplification products are visible). Amplification was obtained for almost all of the tested strains. To determine the size of the PCR fragments was

compared with a commercial size marker (Fermentas). To demine the accurate size the software program gene tools was used, where the band size was determined by the way the cursor was placed on a band and band sizes were estimated.The results from the comparison of the two methods are presented in Table-9.

Table-9: Sizes for PCR fragments for eight different minisatellites amplified from the Proficiency panel, including reference strains as determined by both capillary electrophoresis (CE) and Gel electrophoresis separation. CE Alleles in base pair (bp), Gel repeats: number of repeats based on gel electrophoresis as described in Pourcel *et al.* (2007) The number of repeats in alleles was estimated by subtracting the invariable flanking region from the size and then dividing by the repeat unit length.

Marker	**Lpms1b**			**Lpms3**			**Lpms 13**			**Lpms 17**		
Repeat (bp),	45			96			24			39		
flank(bp)[1]	205			173			164			200		
Strain	Expected[2]	CE repets	Gel repets	Expected[2]	CE repets	Gel repets	Expected[2]	CE repets	Gel repets	Expected[2]	CE repets	Gel repets
Phil_	8	7.7	8.26	8	7.9	8.9	11	11.0	11.3	2	2.1	2.25
Paris_	7	6.8	7.4	7	6.9	7.89	10	10.0	9.9	2	2.1	2.43
Lens_	6	5.8	6.42	7	6.9	7.77	4	4.0	3.7	2	2.1	2.56
EUL 025	9	8.7	9.6	8	7.9	9.65	16	16.1	16.5	2	2.1	3.67
EUL 048	7.5	7.1	7.97	7	6.9	8	12	12.1	12	2	2.1	1.5
EUL 056	7.5	7.1	7.75	7	6.9	8	12	12.1	12	2	2.1	DA
EUL 121	7	6.8	6.84	8	7.9	9	9	9.0	9	2	2.1	2.5
EUL 137	9	8.7	9.8	8	7.9	9.4	9	9.0	9	2	2.1	2.1
EUL 153	8	7.7	8.13	8	7.9	9.18	11	11.0	11.9	1.5	1.5	1.51
EUL 154	8	7.7	7.94	8	7.9	9.0	11	11.0	11.8	1.5	1.5	1.58
EUL 155	8	7.7	7.86	8	7.9	9.33	11	11.0	11.5	1.5	1.5	1.61
EUL 156	9.5	9.0	9.26	8	7.9	9.2	8	8.0	8	2	2.1	2.2
EUL 157	7	6.8	6.8	7	6.9	7.88	10	10.0	10.7	2	2.1	2.15

Marker	**Lpms 19b**			**Lpms 33**			**Lpms 34**			**Lpms35**		
Repeat (bp),	21			125			125			18		
flank(bp)[1]	89			102			84			148		
Strain	Expected[2]	CE repets	Gel repets	Expected[2]	CE repets	Gel repets	Expected[2]	CE repets	Gel repets	Expected[2]	CE repets	Gel repets
Phil	4	3.9	4	1	1.0	1	1	1.0	DA	3	2.9	3
Paris	4	3.9	3.8	4	4.0	4	2	2.0	DA	17	16.7	16.8
Lens_	4	3.9	4	2	2.0	2	3	3.0	DA	23	22.6	22.3
EUL 025	4	3.9	1.68	2	2.0	3.62	2	2.0	DA	16	15.8	15.9
EUL 048	4	3.9	3.7	3	3.0	3	1	1.0	4.52	17	16.7	DA
EUL 056	4	3.9	3.8	3	3.0	3	1	1.0	DA	17	16.7	17
EUL 121	5	4.9	5	3	3.0	3	1	1.0	1	12	11.8	12.5
EUL 137	4	3.9	4	4	4.0	DA	1	1.0	DA	24	23.6	23.8
EUL 153	4	3.9	4.4	1	1.0	1	1	1.0	1	3	2.9	3.2
EUL 154	NA	3.9	4	1	1.0	1	1	1.0	1	3	2.9	3.3
EUL 155	NA	3.9	4	1	1.0	1	1	1.0	1	3	2.9	3.3
EUL 156	4	3.9	4	2	2.0	2	2	2.0	2	18	17.7	18.5
EUL 157	4	3.9	4	4	4.0	4	2	2.0	2	17	16.7	17.3

[1] Deduced from the PCR product sizes and number of repeats for strain Philadelphia-1 according to Pourcel *et al.* (2007), Table 3. [2] Taken from Pourcel *et al.* (2007). DA- difficult to analyze. NA- No amplification

3.2.2. Capillary electrophoresis

In general there was good correspondence between the two different methods both giving similar results. Exceptions were for primer set 1b and 3 where almost one repeat difference between capillary and gel electrophoresis results was observed for most of the strains.

By using different fluorescent dyes attached to the individual primers, multiple PCR products were analysed in the same capillary and typed individually. This is a fast and economic way of separating and determining size of various minisatellites in a single run. Using conventional PCR conditions, several peaks with one repeat size difference in length were observed in the capillary

electrophoresis output (Figure 19 Top). Whereas touchdown PCR resulted in only one peak as seen in figure 19 (bottom).

It is possible that one of the primers binds to the repeat sequence. Two peaks were observed per primer set, this is possibly due to the plus A artefact (see figure-18 which illustrates the two peaks one bp difference due to A artifact). Background smears observed on agarose gels from running PCR products produced using conventional PCR resulted in many peaks in the capillary electrophoresis with one repeat size difference in length (noticable in Figure 19 (top)) . To reduce these effects the PCR programs were optimized for each primer set. Touchdown PCR was used to reduce the background bands, when one one sharp band was observed on the agarose gel then generally one peak was observed in CE (figure-19 (bottom) can see one peak – touchdown result).

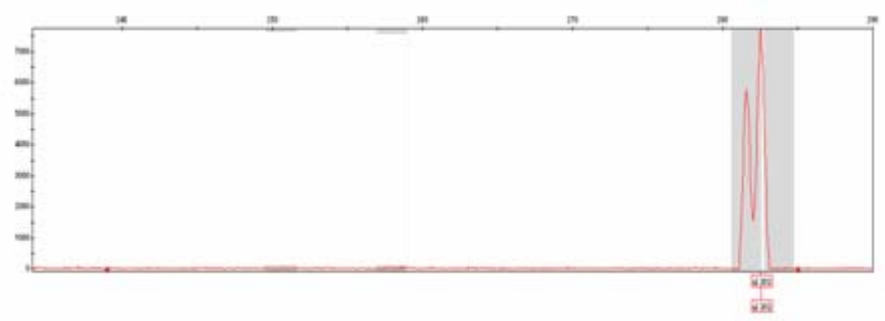

FIGURE 18. Plus A artefact

Chromatogram for strain- 003, marker Lpms- 17 is an example for two peaks that are sometimes observed in capillary electrophoresis separation.

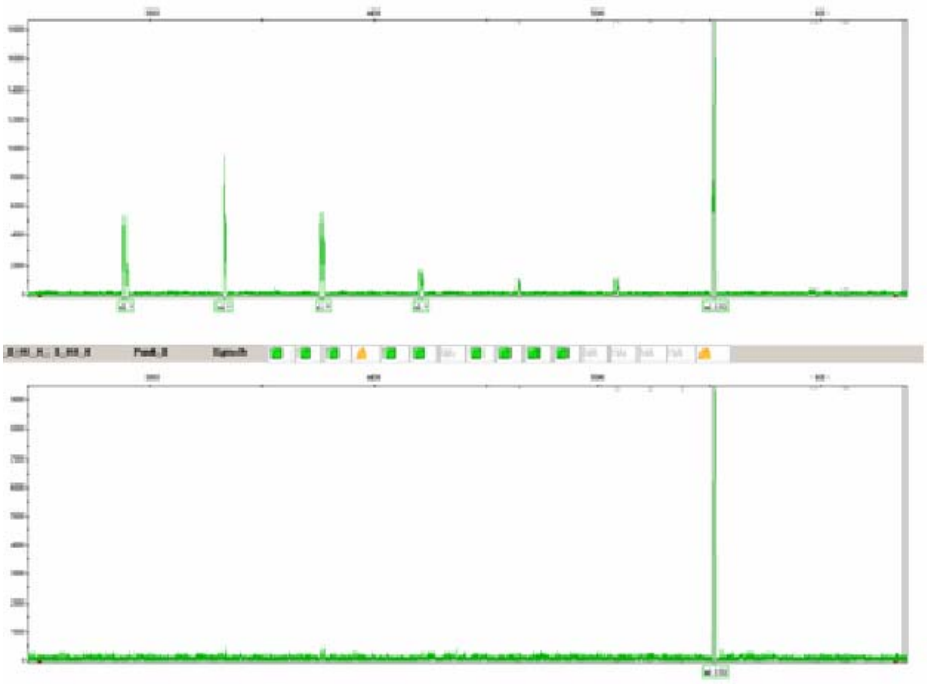

Figure 19. Effect of touchdown PCR

Chromatograms for Strain-003, Lpms-1b.Top: conventional PCR. Bottom: touchdown PCR. The conventional PCR resulted in many peaks. The touchdown PCR resulted in only one peak

3.2.3. Size variation

Upon comparison of the fragment size observed by capillary electrophoresis with theoretical size reported in Pourcel *et al.* (2007) it was noted that in some cases the observed fragment size was longer and in other cases shorter than expected.

For example, when comparing the fragment size of the reference strains *Legionella pneumophila* Philadelphia, Paris and Lens for primer sets Lpms 1b and Lpms 3, large differences were noted between CE and theoretical size

(see Table 10). For primer set Lpms17 a constant size difference of four base pairs was noted for all three strains.

These size differences differ according to the size of the fragment lengths. For example all three strains of *L. pneumophila*, Philadelphia, Paris and Lens, gave 173bp long fragment for primer set Lpms 19. In this situation a 3bp difference was noted. For primer set Lpms 33, for Philadelphia, a fragment length of 224bp was observed and a difference of 3bp noted. For Paris a fragment length of 599bp observed and a difference of 5bp was noted. For Lens a fragment of 348bp was observed and a difference of 4bp was noted. For a particular primer set, when the fragment length increases it appears that the difference also increases, expect for Lpms-34 primer set for strain Paris (see Table 10.).The size variations noted were repeatable.

Table-10 Comparison of fragment length obtained by capillary electrophoresis and expected value obtained by sequencing (from Pourcel *et al.*(2007).The strains used to compare are *Legionella pneumophila* Philadelphia, Paris and Lens.

Primer name	Philadelphia			Paris			Lens		
	E. size[1]	S. p. by CE[2]	difference	E. size[1]	S. p. by CE[2]	difference	E. size[1]	S. p. by CE[2]	difference
Lpms-1b	565	553	-12	520	509	-11	475	465	-10
Lpms-3	941	932	-9	845	837	-8	845	837	-8
Lpms-13	428	429	+1	404	405	+1	260	259	-1
Lpms-17	278	282	+4	278	282	+4	278	282	+4
Lpms-19	173	170	-3	173	170	-3	173	170	-3
Lpms-33	227	224	-3	604	599	-5	352	348	-4
Lpms-34	209	205	-4	334	331	-3	460	454	-6
Lpms-35	202	201	-1	454	449	-5	562	555	-7

[1]E. size -Expected size. [2]S. p. by CE -Size predicted by CE. + means extra bp observed in CE. -Means less bp observe in CE compare to the expected size.

Pourcel *et al.* (2007) in their paper explained that for training purposes and to establish reproducibility of the method the proficiency panel was analysed by three independent laboratories. They reported that no amplification was observed in all three laboratories for Lpms19b in EUL 154 and EUL 155. However, amplification for Lpms19b in both EUL 154 (170 bp) and EUL 155 (170 bp) was obtained during this study. To confirm that the amplification products were Lpms19b repeat sequences, the corresponding PCR products were sequenced. The resulting sequences were aligned using the BLAST tool of the

National Centre for Biotechnology Information (NCBI). BLAST against all known sequences in the database returned a best hit for *L. pneumophila* Strain – Philadelphia (Figure-20 and 21).

```
EUL-154 (019)   CTATCAGAAGGAGGCGATACTTTGAATACAACTGAAATACCAGAACAGCCAATTGAGTAT
L.p-Phil        CTATCAGAAGGAGGCGATACTTTGAATACAACTGAAATACCAGAACAGCCAATTGAGTAT

EUL-154 (019)   CCAGAGGAGCCATCAGAGTACCCTGAACAACCTTTGGAGTACCCTGATAATCCAGAGCCT
L.p-Phil        CCAGAGGAGCCATCAGAGTACCCTGAACAACCTTTGGAGTACCCTGATAATCCAGAGCCT

EUL-154 (019)   CTGGAGCCTGGTTATCCCGAGTTGCCTGAGCCTGAGCCGAGTCAAAC
L.p-Phil        CTGGAGCCTGGTTATCCCGAGTTGCCTGAGCCTGAGCCGAGTCAAAC
```

Figure 20. Lpms19 sequence for strain EUL-154(019)

Alignment of Lpms-19b PCR product for strain EUL-154(019) with the *Legionella pneumophila* Strain - Philadelphia Lpms19b sequence region

```
EUL-155(022)   GAACTATCAGAAGGAGGCGATACTTTGAATACAACTGAAATACCAGAACAGCCAATTGAG
 L.p-Phil      GAACTATCAGAAGGAGGCGATACTTTGAATACAACTGAAATACCAGAACAGCCAATTGAG

EUL-155(022)   TATCCAGAGGAGCCATCAGAGTACCCTGAACAACCTTTGGAGTACCCTGATAATCCAGAG
L.p-Phil       TATCCAGAGGAGCCATCAGAGTACCCTGAACAACCTTTGGAGTACCCTGATAATCCAGAG

EUL-155(022)   CCTCTGGAGCCTGGTTATCCCGAGTTGCCTGAGCCTGAGCCGAGTCAAAC
L.p-Phil       CCTCTGGAGCCTGGTTATCCCGAGTTGCCTGAGCCTGAGCCGAGTCAAAC
```

Figure 21: **Lpms19b sequence for strain EUL_155 (022)**

Alignment of Lpms-19b PCR products for strain EUL-155(022) with *Legionella pneumophila* Strain - Philadelphia Lpms19b sequence region

3.2.4. Additional alleles for marker Lpms-33

By using CE analysis we were able to observe two additional alleles for marker Lpms33 in the proficiency panel, see Table-11.

Table-11: MLVA analysis by capillary electrophoresis identifies additional alleles for marker Lpms-33

strains	CE –bp for Lpms 33	Number of repeats(Taken from Pourcel et al. 2007)
Paris	599bp	4 repeats
EUL137	597bp	4 repeats
EUL157	599bp	4 repeats
Lens	348bp	2 repeats
EUL- 025	349bp	2 repeats
EUL- 156	348bp	2 repeats

3.3 Genotyping of 24 *Legionella* strains from the Norwegian Institute of Public Health

MLVA analysis was performed on 24 strains of *Legionella* from the Norwegian Institute of Public Health. Sizes for PCR fragments for eight different minisatellites on all tested strains as determined by capillary electrophoresis separation are listed in Table-14

Some of the strains tested from the Norwegian Institute of Public Health were not *L. pneumophila* but belonged to other *Legionella* species, for example *L. parisiensis*, *L. londiniensis*, *L. micdadei* and *L. anisa.* Surprisingly it was found that PCR using the MLVA markers on these strains resulted in amplification products from these other Legionella species (see Table-14). This is an unexpected result as Pourcel *et al.* (2007) declared that the MLVA8 typing scheme cannot be used on other *Legionella* species for genotyping and they

observed no amplification when primers were tested on non *L.pneumophilla* species.

To investigate whether these PCR products were tandem repeats of the MLVA minisatellites, some of the amplification products obtained from these non-*pneumophila Legionella* species were sequenced (Lpms-13 and 19 PCR products of Strains 000,003,004,005,006, 007,008 and 009.) Sequencing of the Lpms-13 and -19 PCR products was successful. The resulting sequences were aligned using ClustalW (see figure-22 and 23) to the corresponding sequence regions of the three completely sequenced *L. pneumophila* genomes, *L. pneumophila* str. Philadelphia, *L. pneumophila* str. Paris and *L. pneumophila* str. Lens. (Note that these genomes were originally used to design the PCR primers used in this study, see Pourcel *et al.* 2007). All known sequences in the database retuned best hit for *Legionella pneumophila* except the Lpms-13 PCR product for strain 9 (*L. anisa*). A Blast search against all known sequences returned no hit for this sequence.

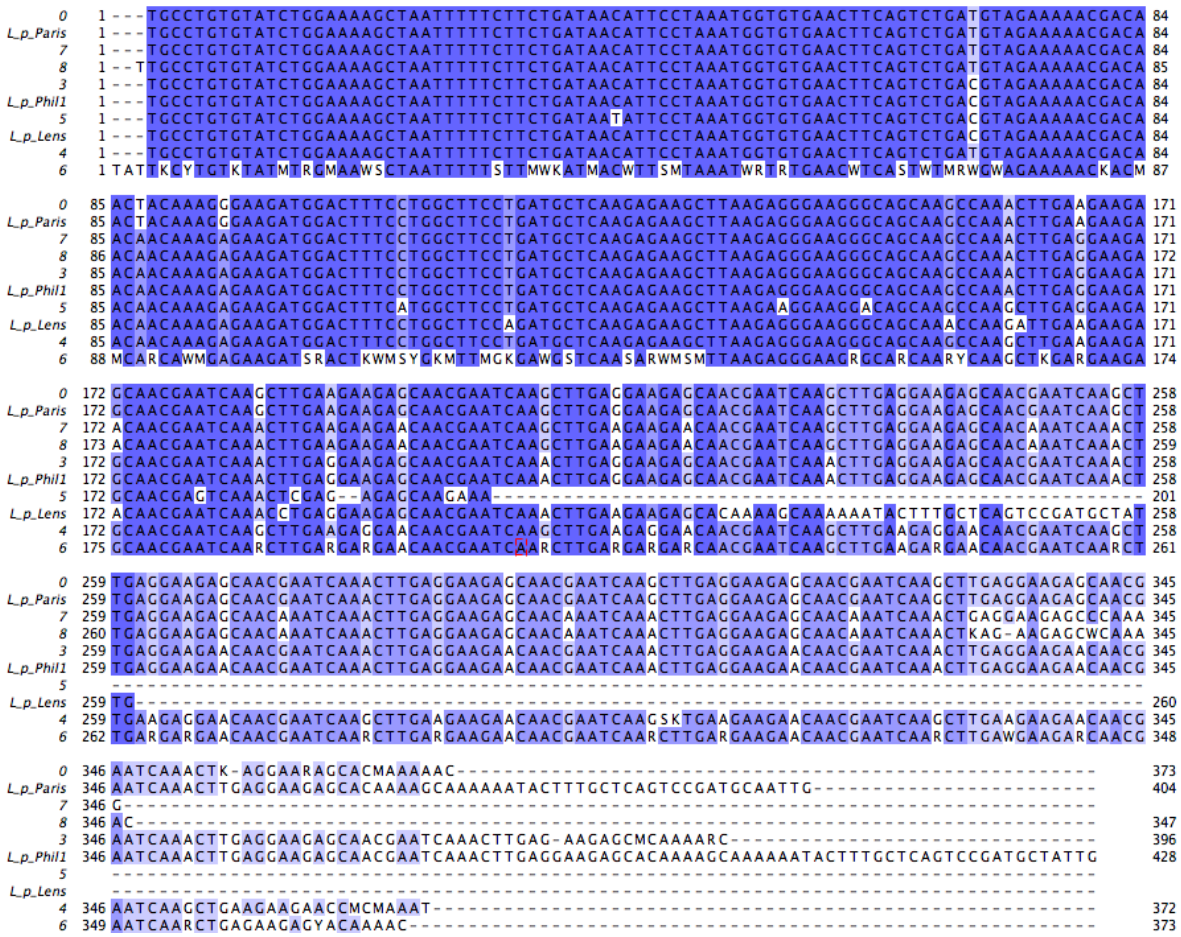

Figure 22. Alignment of Lpms-13 PCR products (strains-0, 3, 4, 5, 6, 7, 8) with the Lpms-13 repeat regions of the reference strains (Philadelphia, Paris and Lens). 100% match-dark blue, Mismatches –no colour. In this figure strain 0, 3, 5, 4, 6, 7 and 8 refers to strain 000, 003, 004, 005, 006, 007 and 008 as detailed in table 4.

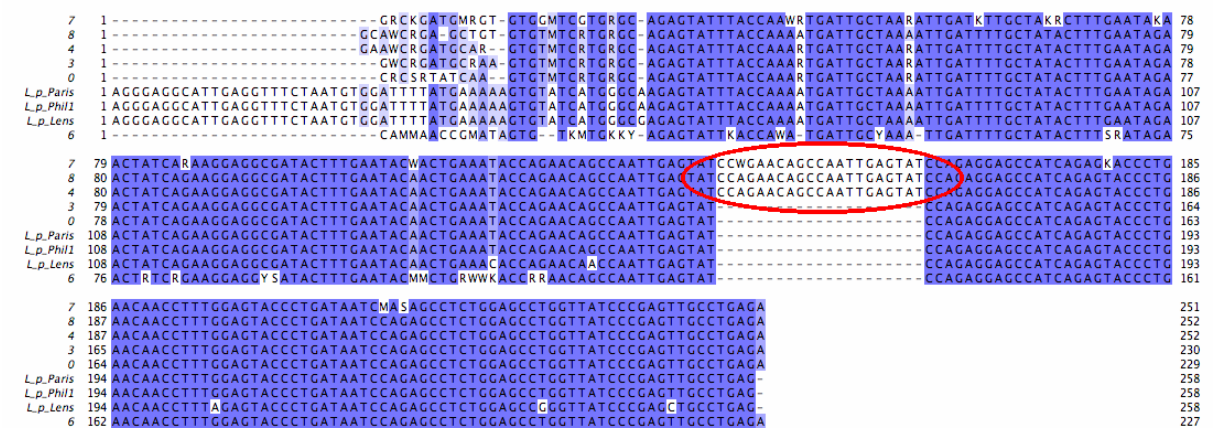

Figure 23. Alignment of Lpms-19b (strains-0, 3, 4, 5, 6, 7, 8) PCR products with the Lpms-19b repeat regions of the reference strains (Philadelphia, Paris, and Lens)

100% match-dark blue, Mismatches –no colour.Note the insertion in three sequences for Lpms-19 (red circle); this is exactly one repeat of the (21 basepair) consensus tandem repeat for this minisatellite. In this figure strain 0, 3, 5, 4, 6, 7 and 8 refers to strain 000, 003, 004, 005, 006, 007 and 008 as detailed in table 4.

3.3.1. Identifying bacteria to species level

The *mip* gene and 16s rDNA are genes that show conserved species-specific polymorphism. These two gene sequences are normally used to determine at the species level.

The *mip* (Macrophage infectivity potentiator) gene is a surface protein that is required for optimal infection of macrophages. Most of the *Legionella* species have a *mip* gene in their genome (Cianciotto and Fields 1992).The *mip* gene was sequenced for the strains analysed in this study. *mip* gene sequencing was employed to identify these bacteria to species level. Sequencing of Mip gene for all strains analysed in ths study confirmed that some of the strains belongs to other *Legionella* species such as *L.parisiensis*, *L.londiniensis*, *Tatlockia micdadei* and *L.anisa.* See tabel 14.

3.4 Comparison of the results for 14 strains obtained from different environments.

The 14 (024-037) of the 24 strains from the Norwegian Institute of Public Health were identified by 16s rDNA analysis and serogroups were determined for some by antibody assay (biochemistry method) by the bacteriology department of the Norwegian Institute of Public Health.

By using the MLVA method only *L. pneumophila* included within this blind panel of *Legionella spp*. were identified and genotyped. MLVA only allows typing of strains of *Legionella pneumophila* and so those bacteria included within the panel not belonging to *L. pneumophila* were characterised using alternative methods(by m*ip* sequencing) (see Table-14)

By comparing the results obtained by MLVA and 16s DNA analysis and *mip* gene sequencing and antibody assay (Table-12) it was noted that strains with identical serotypes as determined by the antibody assay can have different genotypes. *Mip* and 16S rDNA sequencing confirmed that strains 033, 034 and 035 were all *L. pneumophila* and accordingly to antibody analysis were all serogroup 1. However, MLVA analysis revealed that they were two different strains (they had the same genotype as Lens, Paris and Paris respectively). The difference in genotype was detected by MLVA but not *mip* or 16s analysis.

Table 12. Comparison of the results obtained from 16s rDNA and m*ip* sequencing, MLVA analysis and antibody assay for 14 *Legionella spp*..

	Serogroup and identiy of Legionella spp. based on 16s rDNA gene sequencing		**Identiy based on mip gene sequence analysis and strain allocation by MLVA analysis**	
Uio code	**Results based on antibody assay**	**Results based on 16s DNA sequencing**	**Results based on Mip sequencing**	**Genotype based on MLVA**
0 24	Serogroup *	*Legionella erythra*	NA	-NA
025	Serogroup *	*Legionella oakridgensis*	*L. oakridgensis*	- ND
026	Serogroup-1	*Legion.pneumophila*	*L.pneumophila. str. Ph1*	3
027	Serogroup *	*Legionella anisa*	*L. anisa*	- ND
028	Serogroup 2-14	*L. pneumophila*	*L.pneumophila. str.P1*	N.g of L.p.
029	Serogroup*	*L. pneumophila*	*L .pneumophila. str.Ph1*	N.g of L.p
030	Serogroup *	*Legion. lond./naut. Skriv full navn*	*L.londiniensis*	- ND
031	Serogroup -1	*L. pneumophila*	*L.Pneumophila*	3
032	Serogroup *	*Legionella anisa*	*L anisa*	- ND
033	Serogroup -1	*L. pneumophila*	*L. pneumophila. str.Ph*	3
034	Serogroup -1	*L. pneumophila*	*L.pneumophila.*	2
035	Serogroup -1	*L. pneumophila*	*L. pneumophila.str.Ph1*	2
036	Serogroup *	*L. pneumophila*	*L.pneumophila*	N.g of L.p.
037	Serogroup -1	*L. pneumophila*	*L. pneumophila. str. Ph1*	3

L. pneumophila : Legionella pneumophila. 2- Genotype 2 is the same as *paris.* 3- Genotype 3 is the same as *Lens.* NA- No amplification was observed. ND- not detectable. N.g of l.p: New genotype found in *L. Pneumophila.* *: serogroup not determined

Of the total collection of 24 strains from the Norwegian Institute of Public Health, 14 strains were identified as *L. pneumophila* within these, five genotypes were observed in the ten strains of proficiency panel and six new genotypes were observed in the remaining strains.

For strain 024 no amplification was observed for any primer sets (MLVA nor *mip*).

3.5 Location of the repeats in the genome.

Blast searches were used for a database similarity search for these repeats sequences in the NCBI databases. It was found that some of the Lpms repeats are present in between two genes while others are located in-frame in protein coding sequences. Results are displayed in Table 13.

Table 13. Location of the Lpms repeats in the genome

Primer set	Gene	Product
Lpms 1b	In frame within a protein coding region	hypothetical protein
Lpms 3	In frame within a protein coding region	interaptin
Lpms 13	In frame within a protein coding region	coiled-coil-containing protein
Lpms 17	In frame within a protein coding region	hypothetical protein
Lpms 19	inbetween two genes	oxidoreductase dehydrogenase, short indole-3-glycerol phosphate synthase anthranilate phosphoribosyltransferase
Lpms 31	In frame within a protein coding region	tail fiber protein
Lpms 33	inbetween two genes	hypothetical protein Glu/Leu/Phe/Val dehydrogenase
Lpms 34	In frame within a protein coding region	transmembrane protein O-sialoglycoprotein endopeptidase
Lpms 35	In frame within a protein coding region	transmembrane Tfp pilus assembly protein
Lpms 37	inbetween two genes	Nucleotide Metabolism, Replication and DNA polymerase III subunits gamma

Table- 14: Shows fragment sizes for 37 strains determined by capillary electrophoresis and mip sequence result for 37 strains.

Unique genotypes are indicated by the number in the genotype colom.

A: Size for PCR fragments for eight different minisatellites amplified from the reference strains as determined by capillary electrophoresis. CE alleles in base pair

B: Size for PCR fragments for eight different minisatellites amplified from the proficiency panal as determined by capillary electrophoresis. CE alleles in base pair

C: Size for PCR fragments for eight different minisatellites amplified from the24 strains from the Norwegian Institute of Public Health referance strains as determined by capillary electrophoresis. CE alleles in base pair

A

Uio code	Results based on mip	EUL code	Lpms1b	Lpms3	Lpms 13	Lpms 17	Lpms 19	Lpms 33	Lpms 34	Lpms35	genotype
013	*L.pneumophila*	Phil_	553	932	429	282	170	224	205	201	1(PH)
012	*L.pneumophila*	Paris_	509	837	405	282	170	599[d]	331	449	2(P)
011	*L.pneumophila*	Lens_	465	837	259	282	170	348[c]	454	555	3(L)

PH- str.*Philadelphia*
P-str.*paris*
L- Str.*Lens*

B

Uio code	Results based on mip	EUL code	Lpms1b	Lpms3	Lpms 13	Lpms 17	Lpms 19	Lpms 33	Lpms 34	Lpms35	genotype
014	*L.pneumophila*	EUL 048	524	837	454	282	170	473	205	449	4-
015	*L.pneumophila*	EUL 025	597	932	551	282	170	349[c]	331	432	5
016	*L.pneumophila*	EUL 056	524	837	454	282	170	473	205	449	4-EUL048
017	*L.pneumophila*	EUL 137	597	932	381	282	170	597[d]	205	573	6
018	*L.pneumophila*	EUL 121	509	932	381	282	191	473	205	361	7
019	*L.pneumophila*	EUL 154	553	932	429	257	170	224	205	201	1(PH)
020	*L.pneumophila*	EUL 153	553	932	429	257	170	224	205	201	1(PH)
021	*L.pneumophila*	EUL 157	509	837	405	282	170	599[d]	331	449	2(P)
022		EUL 155	553	932	429	257	170	224	205	201	1(PH)
023	*L.pneumophila*	EUL 156	611	932	357	282	170	348[c]	331	467	8

C

Uio code	Results based on mip	EUL code	Lpms1b	Lpms3	Lpms 13	Lpms 17	Lpms 19	Lpms 33	Lpms 34	Lpms35	genotype
000	*L.pneumophila*		509	837	403	282	170	224	329	449	9
001	*L.parisiensis]*		509	837	405	282	191	348	---	509	-
002	*L.londiniensis]*		----	932		282	189	348	----	255	-
003	*L.pneumophila.ph*		552	932	429	282	170	224	204	200	10
004	*L.pneumophila.Ph*		553	932	405	191	191	596	205	343	11
005	*L.micdadei*		539.	906	429	191	191	214	205	343	-
006	*L.micdadei]*				429	282	191		331		-
007	*L.pneumophila.ph*		552	932	381	282	190	348	329	255	12
008	*L.pneumophila.ph*		552	932	381	282	190	348	329	255	12
009	*L. anisa]*				429	282	170	599	331	449	-
024			---	----	---	257	-----	----	-------	----	-
025	*L. oakridgensis*		----	----	-----	----	170	----	------	----	-
026	*L.pneumophila. Ph*		465	837	259	282	170	348	454	555	3(L)
027	*L. anisa*		---	----	----	--	170	-----	------	---	-
028	*L. pneumophila.Ph*		553	932	381	282	170	348	454	348	13
029	*L .pneumophila.Ph*		553	932	381	282	191	348-	331	255	14
030	*L. londiniensis*		----	----	-----	----	170	-----	------	---	-
031	*L.Pneumophila*		465	837	259	282	170	348	454	555	3(L)
032	*L anisa*		------	932	----	----	------		------	---	-
033	*L. pneumophila. str.Ph*		465	837	259	282	170	348	456	555	3(L)
034	*L. pneumophila*		509	837	405	282	170	599	331	449	2(P)
035	*L. pneumophila.Ph*		509	837	405	282	170	599	331	449	2(P)
036	*L .pneumophila*		553	932	381	282	191	348	331	255	9
037	*L. pneumophila.Ph*		465	837	259	282	170	348	456	555	3(L)

DISCUSSION AND CONCLUSION

In this study an existing high resolution genetic typing method known as Multilocus variable number tandem repeat analysis (MLVA) was adapted from agarose gel based to capillary based electrophoresis (CE) analysis of products. Subsequently this method was applied to a number of environmentally and clinically isolated *Legionella pnemophila*. To the best of our knowledge this is the first time that capillary electrophoresis has been used as part of the MLVA analysis of *L. pneumophila*.

MLVA is based on minisatellite analysis. Recently micro and minisatellites have become powerful genetic tools for epidemiology as well as in forensic medicine. It has been shown that minisatellites are well suited genetic makers for molecular analysis. Minisatellites are tandemly repeated units, a type of fragment length variation, caused by slippage of the DNA enzyme during DNA replication (Lindstedt 2005). The length differences are identified with the aid of gel or capillary electrophoresis systems. Genotyping is carried out using software such as Gene Mapper (Applied Biosystems). Based on the fragment length difference calculated, differences between strains can be seen. Since repetitive sequences, such as minisatellites, have a high polymorphism in the fragment lengths and a high mutation rate, differences between closely related strains are detectable.

Capillary electrophoresis is a very sensitive and high resolution technique not least aided by detection of products by laser analysis. In order to adapt the MLVA method for capillary electrophoresis analysis it was necessary to optimise the polymerase chain reaction (PCR) to obtain correct product and concentration. One negative effect of a non optimal PCR reaction is the occurrence of non-specific amplification products which during MLVA analysis by CE appear as small peaks. These small peaks make the chromatogram 'noisey' and appear in addition to the required peak (s). Therefore to improve the efficiency of minisatellite analysis, reaction conditions had to be optimized to avoid nonspecific amplification products such as non specific primer annealing. PCR optimization was performed by modifying reaction conditions including the annealing temperature, concentration of $MgCl_2$, PCR reaction volume and the DNA polymerase enzyme source.

$MgCl_2$ concentrations are known to influence the rigidity of the interaction between the primer and the template DNA. It is therefore an important parameter to adjust during the PCR. Adjusting the concentration is relatively simple. As detailed in the results section many different concentrations were tested until an optimal concentration was reached.

A concentration of 3.5 $MgCl_2$ was chosen for all PCR reactions within our experiments. The reason for selecting this concentration was that when low concentrations were used no amplification products were observed (as determined from agarose gel electrophoresis of the products). When

concentrations over 3.5 $MgCl_2$ were used many bands were observed by agarose gel analysis of the PCR products. It was also investigated whether or not PCR volume could negatively influence the outcome of the reaction. Indeed increasing the reaction volume led to a stronger PCR product as determined by agarose gel analysis. The reason why volume gave more amplification is difficult to explain because the DNA concentration that was used was constant. The DNA extract was from a pure culture so the possibility of the dilution of PCR inhibitors using a higher reaction volume can be discounted. One conclusion that can be drawn from this result is that on some occasions increased volume also can give more amplification product.

An important component in any PCR reaction is DNA polymerase. Within this study 2 DNA polymerases were compared (Go taq and Hot Gold Star). When Go taq DNA polymerase was used in the experiments, background bands were observed. This could be explained by the fact that Go taq DNA polymerase enzyme starts to react at room temperature. When the DNA polymerase reacts at room temperature the primers may anneal at sequences that may only have a few mismatches to the target sequence and therefore result in non-specific amplifications.

When Hot Gold Star DNA polymerase was tested there was a dramatic drop in non-specific amplification products. An explanation for this could be that the Hot Gold Star enzyme does not react at room temperature but has an activation temperature of 95^0c (known as ‘hotstart’). Even after optimizations in a few cases background bands were observed. That means in normal PCR

conditions, background smears were observed in the gel picture and many peaks in the capillary electrophoresis with one repeat difference in the length.

We know that annealing temperature is one of the most important parameters that need adjustment in the PCR reaction. Also annealing temperature is important in finding and documenting polymorphisms. In order to reduce the non-specific amplification temperatures for PCR programs were optimized individually. In cases where large unspecific products appeared in normal PCR, Touchdown PCR was attempted. Touchdown PCR is a method which reduces non-specific amplification, beginning annealing with a higher than optimal temperature and decreasing that temperature steadily to a predetermined temperature. As observed in this project Touchdown PCR dramatically reduced non-specific amplifications. In many cases one band was observed on the agarose gel and one peak detected in the capillary electrophoresis as expected. One primer set (Lpms-17) did cause problems even after touchdown PCR and so gradient PCR was used to determine the optimal annealing temperature. Gradient PCR consists of a number of samples which are all the same being subjected to slightly different annealing temperatures within the one PCR and is carried out in a PCR machine that has a block capable of having different temperatures in parallel wells at the same time. In this experiment some strains gave amplifications at low temperature and other strains at high temperature. This observation shows that annealing temperatures influences amplification, so it was decided to use both low and high temperature for all

strains. To find the reason for the problem that was experienced for primer set-17, the sequences of the flanking regions were examined on strains Philadelphia, Paris and Lens by using sequence analysis and bioinformatics tools. Only within the Lens strain was there a difference noted in the annealing region of the forward primer. Based on the results it was concluded that sequence variation is present within the strains *L. pneumophila* in the region where the forward primer for Lpms17 binds and that this could have been the reason why primer set Lpms-17 gave problems.

After optimizing the PCR conditions, it was necessary to ensure that the MLVA allele assignments inferred from agarose gel electrophoresis were possible to transfer to capillary electrophoresis. For this part the proficiency panel strains used by Pourcel *et al.* (2007) were selected. These same strains were chosen as their genotype is known and it is easy to compare between results obtained in this study with their results. Within this study PCR amplification (using the optimized protocol) for 13 strains was performed with the MLVA 8 primers. Part of the proficiency panel including reference strains - Philadelphia, Paris, Lens were analyzed using agarose gel electrophoresis followed by capillary electrophoresis. The repeatability of this method was examined by comparing expected values (from Pourcel *et al.* 2007) with fragment sizes obtained by both capillary and gel electrophoresis separation. The purpose of this experiment was to ensure that it was possible to transfer the MLVA allele assignments inferred from agarose gel electrophoresis to capillary electrophoresis (as mentioned above), and also to make certain

that these two methods used to determine the fragment length are comparable. Values previously estimated by agarose gel by Pourcel *et al.* 2007 were compared with fragment sizes obtained by both capillary and gel electrophoresis separation within this study.

In general there was good correspondence between the two different methods both giving similar results. Some of the fragment sizes obtained by gel electrophoresis separation (in this study) were too large when comparing with the expected size. This may be due to the way the cursor was placed on a band and band size was estimated (within the gel quantification analysis and software). Due to DNA overloading dense bands will result in the gel picture, when the cursor is placed in the middle of a dense band, the observed size will be not accurate. Another explanation could be due to the gel pore size or may be due to differences during migration between labelled PCR fragments and the ladder used.

The development of capillary electrophoresis has led to a number of benefits, speed, high separation efficiency, low sample consumption and the ability to analyze multiple PCR products in the same capillary. Its high capacity with the use of 96 well plates allows 96 samples to be analyzed simultaneously. Other advantages include rapid separation of the DNA amplification products in an automated fashion and quantitative/ semi quantitative results. All these advantages result in time saving and cost reduction and most importantly accurate and reliable results. It is possible to run differently labelled amplification products (which show as different colours on the output)

together and type individually. Additionally, as fluorescence is laser detected it makes this method highly accurate and repeatable. This is a fast and economic way of separating and determining size of various minisatellites in a single run (Robert, Bouilloux et al. 1991). The assay was highly repeatable as tested by performing duplicate typing. The accuracy of the data was determined by comparison of observed PCR fragment size to the exact values calculated by direct sequencing and previously estimated by agarose gel by Pourcel *et al.* 2007.

Upon careful examination of the results acquired from the CE it was noted that the fragment size determined after CE separation did not match the length determined by sequencing. Thus, the fragment size obtained from CE is longer or shorter when compared with the expected value. Even though size variation was observed the variation was repeatable. An explanation to this variation may be due to the different fluorescent dyes accompanying the fragments during migration or sequence specific migration behaviour may be the alternative explanation. Another observation was that the variation in most cases increases with increasing fragment length. It is possible that sequence specific migration behaviour may be an explanation for this observation. This type of variation differences in actual size and observed using CE has been noted in other studies using CE based MLVA typing example in the typing of *Bacillus anthracis* (Lista, Faggioni et al. 2006).

In some cases two peaks were observed in the output with one basepair difference. The reason for the additional peak may be due to the DNA

polymerase which adds an extra A nucleotide to the 3`end of PCR products after extinction of the template. The base T (tymidine) at the end of the template can cause problems; when DNA polymerase copies this T it has to add only C to the growing PCR product but it adds both C and A . This is the reason why an extra base occurs and produces the plus A peak. (Brownstein, Carpten et al. 1996). This sort of enzyme activity can cause problems in allele calling during genotyping. This is primer specific activity and can be resolved by modifying or adding extra bases as a pig tail at the beginning of the primers (Magnuson, Ally et al. 1996). As this is a reoccurring problem which has a satisfactory explanation, it can be ignored and the plus A peak can be manually deleted from the chromatogram.

Sizing PCR products obtained by capillary electrophoresis is more sensitive than gel electrophoresis. By using CE analysis, two additional alleles were observed for marker Lpms33 in the proficiency panel. This demonstrates that the CE based assay is more sensitivity than the gel based genotyping assay. This occurrence shows that fragment length based genotyping gives higher resolution accurate to between 1 and a few base pairs. Another indication of the higher discriminatory power of CE over conventional gel electrophoresis was the observation of 2 alleles for Lpms 33 as opposed to one by Pourcel et al. 2007.

Analysing the results from this study I recommend that fragment length based genotyping (based on determining the length of amplified tandem repeat fragments) gives more accurate resolution than repeat based genotyping

(based on the number of tandem repeats for each amplified fragment). Some strains may differ from others by only a few base pair in one marker. This difference can be detected using CE and analysis based on fragment length (number of bp) but not by using a gel based genotyping assay. Furthermore using the repeat based genotyping few (one or two bp) bp differences will be neglected.

Another important finding in this study was the observation of amplification products for Lpms- 19 for strains EUL-154 and EUL-155. This result was in contrast to that of Pourcel *et al.* (2007), where 3 independent laboratories did not obtain a product for this reaction. The result obtained here was confirmed by sequencing which shows that strains EUL-154 and EUL-155 have the identical repetitive sequence as sequences present in the genome of the *L. pneumophila* Strain - Philadelphia. So it is clear that the PCR amplification obtained was indeed not an artefact. It also illustrates the importance of optimization of all PCR reaction conditions as carried out during this study. One reason for this difference could be the high-quality PCR polymerase that was used in this study. Alternatively it may just be indicative of the effort that was put in order to optimise all the PCR reactions individually.

Later in the study this improved MLVA 8 method was applied to 24 strains of both *L. pneumophila* and other Legionella *species*. For other non-*L. pneumophila* species amplification products for some of the MLVA8 markers were obtained. This result was unexpected, since Pourcel and co-workers (2007) did not find any amplification products for non L*. pneumophila* species.

Results from sequencing conformed that these were indeed repetitive sequences similar to the sequences present in the genome of the *L. pneumophila* Strain - Philadelphia. The improved PCR condition or PCR enzyme used may have been a reason for the amplification products obtained.

14 of the 24 strains from the Norwegian Institute of public Health had been previously characterized by 16S rDNA analysis and serogroup typing. When the results obtained by MLVA and 16s rDNA analysis and *mip* gene sequencing were compared there was a high level of concordance between these methods. *mip* gene-Macrophage infectivity potentiator is a surface protein that is required for optimal infection of macrophages. Most of the *Legionella* species have a *mip* gene in their genome (Cianciotto, N. P., and B. S. Fields. 1992). Sequence analaysis of these genes can help determine bacteria to the species level. However, the MLVA method was found to be superior to the other methods, because of its high discrimination at the strain level. Additionally it was observed that identical serogroups can have different genotypes and there is as yet no known way of predicting serogroup by analysis of genetic markers. One advantage of sequence analysis of mip and 16s rDNA genes is that they allow the identification of the species. MLVA can then be used to genotype *L. pneumophila* species and other species can be disregarded.

A number of articles detail the relationship between tandem repeats within protein coding regions and disease. For example in Huntingtons disease

people suffering from this disease have been found to have longer repetitive sequences than healthy individuals. O'Dushlaine, Edwards et al. 2005) (Verkerk, Pieretti et al. 1991; O'Dushlaine, Edwards et al. 2005)

It has also been noted within prokaryotic pathogens such as *Nesseria spp*. and *Bacillus anthracis* that they contain more tandem repeats within protein coding regions than non pathogenic relatives(Jordan, Snyder et al. 2003; Sylvestre, Couture-Tosi et al. 2003). It has been found that variation in the repeat number within the coding region is associated with antigen variation in certain pathogens. (Jordan, Snyder et al. 2003; Sylvestre, Couture-Tosi et al. 2003). *L. pneumophila* has been widely documented as the causative agent of many outbreaks of Legionnaires disease (Aurell, Farge et al. 2005). The reasons why this species is particularly virulent remains unclear.

Upon further investigation it was found that these *L.pneumophila* MLVA 8 repeats are present in between two genes and also in between protein sequences. Blast algorithms were used for a database similarity search for these repeats sequences in the NCBI databases. It was found that these repeat sequences were present in coding regions. It is possible that, this could indicate a role of tandem repeats in virulence but this requires further investigation and comparison with non pathogenic *Legionella* species, these *L.pneumophila* MLVA-8 repeats can also alter the amino acid sequence of the corresponding protein, it has for example been shown that changes in cell wall proteins due to the number of TRs can cause alterations in

immunogenicity, adhesion and pathogenesis. Again this study produces no evidence to corroborate these hypotheses and further research is necessary.

All together 27 *L.pneumophilla* strains were analysed by MLVA typing and 14 different genotypes of *L.pnemophilla* were detected. It is suggested that these data can be maintained as part of a genotype library at the Norwegian Institute of Public Health. Future outbreaks or clinical strains can then be compared to the growing database of MLVA genotypes making links between outbreaks and occurrences of Legionnaires disease easier. This could be a useful tool allowing different researchers access to a growing database of results from clinical and environmental isolates.

In conclusion, MLVA on capillary electrophoresis is the best method to analyse the different strains and genotype the *L. pnemophila.* The current protocol was designed to avoid extensive use of sequencing facilities (as necessary when using MLST analysis) and was very specific. MLVA assay by capillary electrophoresis was capable of high discrimination between the *L. pneumophila* strains. The assay was fast and all the steps were automated. The technique also lends itself to easy portability of results between laboratories, which is a necessary feature in the investigation of disease outbreaks.

References

(1993). "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group." Cell **72**(6): 971-83.

Adair, D. M., P. L. Worsham, et al. (2000). "Diversity in a variable-number tandem repeat from Yersinia pestis." J Clin Microbiol **38**(4): 1516-9.

Atlas, R. M. (1999). "Legionella: from environmental habitats to disease pathology, detection and control." Environ Microbiol **1**(4): 283-93.

Aurell, H., P. Farge, et al. (2005). "Clinical and environmental isolates of Legionella pneumophila serogroup 1 cannot be distinguished by sequence analysis of two surface protein genes and three housekeeping genes." Appl Environ Microbiol **71**(1): 282-9.

Berdal, B. P., C. E. Farshy, et al. (1979). "Detection of Legionella pneumonophila antigen in urine by enzyme-linked immunospecific assay." J Clin Microbiol **9**(5): 575-8.

Blessum, C., J. O. Jeppsson, et al. (1999). "[Capillary electrophoresis: principles and practice in clinical laboratory]." Ann Biol Clin (Paris) **57**(6): 643-57.

Bois, P. R. (2003). "Hypermutable minisatellites, a human affair?" Genomics **81**(4): 349-55.

Brenner, D. J., N. R. Krieg, et al. (2005). Bergey's manual of systematic bacteriology, Springer.

Brown, T. A. (2007). Genomes 3, Garland Science.

Brownstein, M. J., J. D. Carpten, et al. (1996). "Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping." Biotechniques **20**(6): 1004-6, 1008-10.

Chang, C. H., Y. C. Chang, et al. (2007). "VNTRDB: a bacterial variable number tandem repeat locus database." Nucleic Acids Res **35**(Database issue): D416-21.

Cherry, W. B., B. Pittman, et al. (1978). "Detection of Legionnaires disease bacteria by direct immunofluorescent staining." J Clin Microbiol **8**(3): 329-38.

Cianciotto, N. P. and B. S. Fields (1992). "Legionella pneumophila mip gene potentiates intracellular infection of protozoa and human macrophages." Proc Natl Acad Sci U S A **89**(11): 5188-91.

Coletta-Filho, H. D., M. A. Takita, et al. (2001). "Differentiation of strains of Xylella fastidiosa by a variable number of tandem repeat analysis." Appl Environ Microbiol **67**(9): 4091-5.

Farlow, J., K. L. Smith, et al. (2001). "Francisella tularensis strain typing using multiple-locus, variable-number tandem repeat analysis." J Clin Microbiol **39**(9): 3186-92.

Fields, B. S., R. F. Benson, et al. (2002). "Legionella and Legionnaires' disease: 25 years of investigation." Clin Microbiol Rev **15**(3): 506-26.

Fliermans, C. B., W. B. Cherry, et al. (1981). "Ecological distribution of Legionella pneumophila." Appl Environ Microbiol **41**(1): 9-16.

Fraser, D. W., T. R. Tsai, et al. (1977). "Legionnaires' disease: description of an epidemic of pneumonia." N Engl J Med **297**(22): 1189-97.

Frazier, A. R., M. J. Ames, et al. (2000). Capillary electrophoresis for food analysis, RS-C.

Gaia, V., N. K. Fry, et al. (2005). "Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of Legionella pneumophila." J Clin Microbiol **43**(5): 2047-52.

Grimont, P. A., F. Grimont, et al. (1985). "DNA probe specific for Legionella pneumophila." J Clin Microbiol **21**(3): 431-7.

http://gsbs.utmb.edu/microbook/ch040.htm Legionella. W. J. Winn, MedMicro.

Hussong, D., R. R. Colwell, et al. (1987). "Viable legionella pneumophila not detectable by culture on agar media." Biotechnology **5**.

Jantzen, E. and I. Olsen (2002). "Legionella-bakterien, hvor lever den og hvordan gir den sykdom?" Naturen(4): 194-199.

Joly, P., P. A. Falconnet, et al. (2006). "Quantitative real-time Legionella PCR for environmental water samples: data interpretation." Appl Environ Microbiol **72**(4): 2801-8.

Jordan, P., L. A. Snyder, et al. (2003). "Diversity in coding tandem repeats in related Neisseria spp." BMC Microbiol **3**: 23.

Joseph, C. A. (2004). "Legionnaires' disease in Europe 2000-2002." Epidemiol Infect **132**(3): 417-24.

Keim, P., L. B. Price, et al. (2000). "Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis." J Bacteriol **182**(10): 2928-36.

Le Fleche, P., M. Fabre, et al. (2002). "High resolution, on-line identification of strains from the Mycobacterium tuberculosis complex based on tandem repeat typing." BMC Microbiol **2**: 37.

Le Fleche, P., Y. Hauck, et al. (2001). "A tandem repeats database for bacterial genomes: application to the genotyping of Yersinia pestis and Bacillus anthracis." BMC Microbiol **1**: 2.

Liao, J. C., C. C. Li, et al. (2006). "Use of a multilocus variable-number tandem repeat analysis method for molecular subtyping and phylogenetic analysis of Neisseria meningitidis isolates." BMC Microbiol **6**: 44.

Lindstedt, B. A. (2005). "Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria." Electrophoresis **26**(13): 2567-82.

Lindstedt, B. A., E. Heir, et al. (2003). "DNA fingerprinting of Salmonella enterica subsp. enterica serovar typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci." J Clin Microbiol **41**(4): 1469-79.

Lista, F., G. Faggioni, et al. (2006). "Genotyping of Bacillus anthracis strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis." BMC Microbiol **6**: 33.

Liu, Y., M. A. Lee, et al. (2003). "Molecular typing of Salmonella enterica serovar typhi isolates from various countries in Asia by a multiplex PCR assay on variable-number tandem repeats." J Clin Microbiol **41**(9): 4388-94.

Magnuson, V. L., D. S. Ally, et al. (1996). "Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning." Biotechniques **21**(4): 700-9.

Mauchline, W. S., B. W. James, et al. (1994). "Growth temperature reversibly modulates the virulence of Legionella pneumophila." Infect Immun **62**(7): 2995-7.

McDade, J. E., C. C. Shepard, et al. (1977). "Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease." N Engl J Med **297**(22): 1197-203.

Moss, C. W., R. E. Weaver, et al. (1977). "Cellular fatty acid composition of isolates from Legionnaires disease." J Clin Microbiol **6**(2): 140-3.

Nygard, K. (2005). "Update: outbreak of legionnaires disease in Norway traced to air scrubber." Euro Surveill **10**(6): E050609 1.

O'Dushlaine, C. T., R. J. Edwards, et al. (2005). "Tandem repeat copy-number variation in protein-coding regions of human genes." Genome Biol **6**(8): R69.

O'Toole, G., H. B. Kaplan, et al. (2000). "Biofilm formation as microbial development." Annu Rev Microbiol **54**: 49-79.

Par, A., I. Kantor, et al. (1991). "Prevalence of antibody to hepatitis C virus in blood donors, high-risk groups and patients with liver diseases in Hungary. A multicentre study using ABBOTT EIA test and a comparison with an ORTHO ELISA test system." Acta Med Hung **48**(3-4): 167-76.

Parham, P. (2005). The immune system Garland Science.

Pattison, J. R., G. R. N., et al. (2006). A practical guide to clinical bacteriology.

Pourcel, C., Y. Vidgop, et al. (2003). "Characterization of a tandem repeat polymorphism in Legionella pneumophila and its use for genotyping." J Clin Microbiol **41**(5): 1819-26.

Pourcel, C., P. Visca, et al. (2007). "Identification of variable-number tandem-repeat (VNTR) sequences in Legionella pneumophila and development of an optimized multiple-locus VNTR analysis typing scheme." J Clin Microbiol **45**(4): 1190-9.

Robert, F., J. P. Bouilloux, et al. (1991). "[Capillary electrophoresis: principle and applications]." Ann Biol Clin (Paris) **49**(3): 137-48.

Rowbotham, T. J. (1980). "Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae." J Clin Pathol **33**(12): 1179-83.

Sabat, A., J. Krzyszton-Russjan, et al. (2003). "New method for typing Staphylococcus aureus strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates." J Clin Microbiol **41**(4): 1801-4.

Smith, A. and R. J. Nelson (2004). "Capillary electrophoresis of DNA." Curr Protoc Mol Biol **Chapter 2**: Unit 2 8.

States, S. J., L. F. Conley, et al. (1985). "Effects of metals on Legionella pneumophila growth in drinking water plumbing systems." Appl Environ Microbiol **50**(5): 1149-54.

Steinert, M., U. Hentschel, et al. (2002). "Legionella pneumophila: an aquatic microbe goes astray." FEMS Microbiol Rev **26**(2): 149-62.

Stolhaug, A. and K. Bergh (2006). "Identification and differentiation of Legionella pneumophila and Legionella spp. with real-time PCR targeting the 16S rRNA gene and species identification by mip sequencing." Appl Environ Microbiol **72**(9): 6394-8.

Sylvestre, P., E. Couture-Tosi, et al. (2003). "Polymorphism in the collagen-like region of the Bacillus anthracis BclA protein leads to variation in exosporium filament length." J Bacteriol **185**(5): 1555-63.

Tilton, R. C. (1979). "Legionnaires' disease antigen detected by enzyme-linked immunosorbent assay." Ann Intern Med **90**(4): 697-8.

van Belkum, A., W. J. Melchers, et al. (1997). "Outbreak of amoxicillin-resistant Haemophilus influenzae type b: variable number of tandem repeats as novel molecular markers." J Clin Microbiol **35**(6): 1517-20.

Verkerk, A. J., M. Pieretti, et al. (1991). "Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome." Cell **65**(5): 905-14.

Wadowsky, R. M., R. Wolford, et al. (1985). "Effect of temperature, pH, and oxygen level on the multiplication of naturally occurring Legionella pneumophila in potable water." Appl Environ Microbiol **49**(5): 1197-205.

Wilkinson, H. W. and B. J. Brake (1982). "Formalin-killed versus heat-killed Legionella pneumophila serogroup 1 antigen in the indirect immunofluorescence assay for legionellosis." J Clin Microbiol **16**(5): 979-81.

Wullings, B. A. and D. van der Kooij (2006). "Occurrence and genetic diversity of uncultured Legionella spp. in drinking water treated at temperatures below 15 degrees C." Appl Environ Microbiol **72**(1): 157-66.

Wyman, A. R. and R. White (1980). "A highly polymorphic locus in human DNA." Proc Natl Acad Sci U S A **77**(11): 6754-8.

Yazdankhah, S. P., B. A. Lindstedt, et al. (2005). "Use of variable-number tandem repeats to examine genetic diversity of Neisseria meningitidis." J Clin Microbiol **43**(4): 1699-705.

Appendix 1. Manuscript of the accepted article (in Journal of Microbiological Methods) based on part of this study

Multiple-locus variable-number tandem repeat analysis of *Legionella pneumophila* using multi-colored capillary electrophoresis

Alexander J. Nederbragt, Anusha Balasingham, Reidun Sirevåg, Hans Utkilen, Kjetill S. Jakobsen, Mary Anderson-Glenna

PII: S0167-7012(08)00038-9
DOI: doi: 10.1016/j.mimet.2008.02.007
Reference: MIMET 2898

To appear in: *Journal of Microbiological Methods*

Received date: 24 October 2007
Revised date: 6 February 2008
Accepted date: 7 February 2008

Please cite this article as: Nederbragt, Alexander J., Balasingham, Anusha, Sirevåg, Reidun, Utkilen, Hans, Jakobsen, Kjetill S., Anderson-Glenna, Mary, Multiple-locus variable-number tandem repeat analysis of *Legionella pneumophila* using multi-colored capillary electrophoresis, *Journal of Microbiological Methods* (2008), doi: 10.1016/j.mimet.2008.02.007

Multiple-locus variable-number tandem repeat analysis of *Legionella pneumophila* using multi-colored capillary electrophoresis

Alexander J. Nederbragt[1], Anusha Balasingham[2,3], Reidun Sirevåg[3], Hans Utkilen[2], Kjetill S. Jakobsen[1#], Mary Anderson-Glenna[2].

[1] University of Oslo, Department of Biology, Centre for Ecological and Evolutionary Synthesis, P.O. Box 1066 Blindern, 0316 Oslo, Norway

[2] Department of Water Hygiene, Division of Environmental Medicine, Norwegian Institute for Public Health, P.O. Box 4404 Nydalen, N-0403 Oslo, Norway

[3] Department of Molecular Biosciences, University of Oslo, P.O. Box 1066, Blindern, N-0316 Oslo, Norway

[#] Corresponding author:

Kjetill S. Jakobsen

University of Oslo

Department of Biology, Centre for Ecological and Evolutionary Synthesis

P.O. Box 1066 Blindern

0316 Oslo, Norway

Email: k.s.jakobsen@bio.uio.no

Phone: +47-22854602, +47-22854603; Fax +47-22854605

Abstract

Several methods for typing of *Legionella pneumophila* exist, one of which is an 8-locus Variable Number of Tandem Repeats analysis (MLVA). This method is based on separating and sizing amplified VNTR PCR products by agarose gel electrophoresis. In the present work, the existing *L. pneumophila* MLVA-8 assay is adapted to capillary electrophoresis. The assay was multiplexed by using multiple fluorescent labeling dyes and tested on a panel of *L. pneumophila* strains with known genotypes. The results from the capillary electrophoresis-based assay are shown to be equivalent to, and in a few cases more sensitive than, the gel-based genotyping assay. The assay presented here allows for a swift, automated and precise typing of *L. pneumophila* from patient or environmental samples and represents an improvement over the current gel-based method.

Keywords

Legionella; Genotyping; Capillary Electrophoresis, Multiple-Locus Variable-Number Tandem Repeat Analysis (MLVA)

Introduction

Legionella pneumophila, a Gram-negative non-spore-forming rod, is the causative agent of the majority of cases of Legionnaires disease and a milder form, Pontiac fever. *L. pneumophila* is transmitted via microscopic droplets of water as aerosols originating from cooling towers, hot water systems, air-conditioning systems, whirlpool baths etc. In the natural environment a number of free-living amoeba and ciliated protozoa, such as *Acanthamoeba*, *Naegleria* and *Balamuthia*, serve as host for *L. pneumophila* (Borella et al., 2005), but the organism has also been observed to exist within aquatic biofilms (Declerck et al., 2007). Individuals most at risk from infection include those with compromised immune or respiratory systems, heavy smokers and elderly people. Within human lungs *L. pneumophila* are engulfed by the pulmonary macrophages where they multiply and persist intracellularly.

Initially characterized as the causative agent of the 1976 outbreak at a legionnaires conference where 221 people were infected and 34 people died, mortality rates attributed to Legionnaires disease have continued to increase (McDade, 2002). For example in southeast Norway in May 2005, 55 people were infected resulting in 10 deaths. This outbreak was eventually traced to an air scrubber installation (Nygard, 2005). A recent outbreak of Legionnaires disease in a care facility in Toronto, Canada claimed the lives of 23 elderly people. A total of 135 individuals were infected with the disease that was traced to the facilities cooling tower (Henry et al., 2005). At a flower show in the Netherlands in 1999 at least 188 people became infected with Legionnaires disease from a number of sources, including a sprinkler and whirlpool spas, within the exhibition area. Of the 188 infected 32 people died due to the disease (Den Boer et al., 2002).

Due to its widespread occurrence in aquatic systems during outbreak situations it is of the utmost importance to be able to rapidly and precisely trace the source of infection to allow implementation of control measures. Among the several typing methods for

Legionella pneumophila, two of the most important are a Multilocus sequencing typing scheme (Gaia et al., 2005, Ratzow et al., 2007) and a Multi-locus variable number tandem repeat analysis (MLVA) assay (Pourcel et al., 2003, Pourcel et al., 2007). The first assay relies on DNA sequencing of multiple polymorphic DNA segments, while in the latter, strain discrimination is based on size differences due to VNTR repeat length variation.

The MLVA assay consists of 8 VNTR loci (MLVA-8), where tandem repeats (minisatellites) used as markers vary in repeat unit length from 7 to 125 bp, and in repeat number from a few (1-4) up to 32 repeats. The PCR products obtained are in the range from 150 to close to 1000 bp. The assay as described was based on agarose gel electrophoretic separation of PCR products and size calling by gel image analysis software (Pourcel, et al., 2007). The number of repeats was determined based on the estimated sizes and these repeat numbers were used as alleles.

Another approach to determine PCR product lengths in MLVA is automated capillary electrophoresis (CE, van Belkum, 2007). CE has several advantages. First, by using a microtiter plate format, up to 96 samples can be processed in parallel, allowing for high throughput analysis. Second, each sample contains its own size ladder, allowing a more precise determination of PCR product size and eliminating distortions that can occur during agarose gel electrophoresis. Third, CE allows for resolution down to (fractions of) a basepair. Such a high resolution can resolve fragments that would not be separable with agarose gel electrophoresis (Lista et al., 2006). Furthermore, by using different fluorescent dyes, multiple PCR products can be analyzed in the same capillary, reducing cost and time. This multiplexing can either be performed after or during PCR (multiple primer sets per PCR reaction, each with its own fluorescent dye). Sample loading, separation and size calling are automated. Specific software has been developed to assist in data interpretation. The higher accuracy, greater resolution and time savings makes CE an excellent choice for MLVA. In a direct comparison of agarose gel electrophoresis

with capillary electrophoresis for VNTR analysis of *Mycobacterium tuberculosis*, it has been reported that capillary electrophoresis should be used for correct amplicon size measurement (Yokoyama et al., 2006).

The purpose of the present work was to adapt the *L. pneumophila* MLVA-8 assay of Pourcel et al. (2007) to capillary electrophoresis. When the multiplexed method, using capillary electrophoresis for the separation and automated detection of fragments, was employed to a panel of strains with known genotypes, the showed the same accuracy, and displayed in addition higher sensitivity than the gel-based genotyping assay.

Materials and Methods

Bacterial strains used in this study were the same as those used by Pourcel et al. (2007) as part of their proficiency panel and including the reference strains *Legionella pneumophila* Philadelphia-1, Paris and Lens. *L. pneumophila* Philadelphia-1 (NCTC 11192) was obtained form the National Collection of Type Cultures, London, United Kingdom. All other strains were obtained from the European Working Group on Legionella Infections (EWGLI) culture collection having the specific EUL culture collection numbers 146 (strain Paris), 160 (strain Lens), 025, 048, 056, 121, 137 and 153 - 157.

DNA was isolated from pure cultures of *L. pneumophila* that had been cultured on Buffered Charcoal Yeast Extract agar at 37°C in a humidifying chamber. DNA was extracted using a cetyl trimetyl ammonium bromide (CTAB) based extraction (van Soolingen et al., 1999). Briefly, a small amount of cells were resuspended in 400 µl of TE buffer (10 mM Tris- HCL and 1 mM EDTA (pH 8)) in an Eppendorf tube. 50 µl of 1% (w/v) Lysozyme was added and the mixture was incubated at 37°C for 3 hours. 5 µl of 1% (w/v) Proteinase K and 70 µl of 10% SDS were added, mixed and the mixture was incubated for 15 minutes at 65°C, before 100 µl each of 5M NaCl and CTAB were added and incubated for a further 10 minutes. Finally, 750 µl of chloroform/ isoamyl alcohol were added and the mixture was centrifuged for 5 minutes at maximum speed in a 5804R centrifuge (Eppendorf, Hamburg, Germany). The water phase was transferred to a new Eppendorf tube and 450 µl of isopropanol was added and mixed. The DNA was left to precipitate at -20°C overnight. Tubes were centrifuged at max speed for 10 minutes, the supernatant removed and the pellet washed with ice cold 70% ethanol. The pellet was air-dried and DNA resuspended in 100 µl of TE buffer.

VNTR amplification

The PCR primers used are described by Pourcel et al. (2007). Forward primers were labeled with VIC, PET and NED (Applied Biosystems, Chesire, United Kingdom) or FAM (MWG, Ebersberg, Germany), see Table I. Reverse primers were synthesized unlabelled (MWG, Ebersberg, Germany). 10 ng DNA was amplified in a 25 µl reaction mix containing 1 x buffer (Eurogentec, Seraing, Belgium), 3.5 mM $MgCl_2$ (Eurogentec), 0.2 mM dNTP (Saveen&Werner, Kristiansand, Norway), 0.4 µM of each primer and 0.625 U HotGoldStar DNA polymerase (Eurogentec). PCR cycling conditions for markers Lpms (*Legionella pneumophila* minimicrosatellite) 1b, 3, 13 and 34 were 10 min. at 95°C; 20 cycles touchdown PCR [15 sec. at 95°C; 30 sec. at 75°C, with 0.5°C drop in temperature each next cycle; 45 sec. at 72°C]; 15 cycles regular PCR [15 sec. at 95°C; 30 sec. at 65°C; 45 sec. at 72°C]; with a final 7 min. at 72°C. For marker Lpms35, the same program was used, except that the final regular PCR cycling was done for 20 cycles. For Lpms19b and 33, the same program was used as for Lpms1b, except that for the touchdown cycling, the starting temperature was 70°C, and annealing during the regular PCR cycles was at 60°C. For marker Lpms17, the following program was used: 10 min. at 95°C; 40 cycles regular PCR [15 sec. at 95°C; 30 sec. at 46°C; 45 sec. at 72°C]; with a final 7 min. at 72°C; in addition, for this primer set an identical PCR program was used on all DNA samples, however the annealing temperature was set to 57.8 °C instead of 46°C. Amplifications were performed on a DNA Engine Tetrad2 or PTC200 (MJResearch, Watertown, MA, USA).

Capillary electrophoresis

1 µl of PCR products for markers Lpms13, 19b and 34 and 1 µl for each of the two Lpms17 PCRs were pooled and diluted to 100 µl (Panel I). 1 µl of PCRs for markers Lpms1b, 33 and 35 and 3 µl for marker Lpms3 were pooled and also diluted to 100 µl (Panel II). To the wells of a 96 well microtiter plate, 1 µl of a pooled, diluted PCR product

mix was added to 8.8 µl HiDi Formamide (Applied Biosystems, Foster City, CA, USA) and 0.2 µl GeneScan 1200 LIZ Size standard (Applied Biosystems), which contains 73 single-stranded labeled fragments in the range from 20 to 1200 bp. Samples were denatured for 3 min. at 95°C, cooled on ice for at least one minute and the microtiter plate was spun briefly at 500 rpm in an 5810R centrifuge (Eppendorf). Fragment analysis was performed on a 3730xl sequencer (Applied Biosystems) equipped with 50 cm capillaries, using POP-7 polymer, with the recommended running parameters for the GeneScan 1200 LIZ Size standard. The results were analyzed with GeneMapper software, v3.7 (Applied Biosystems), using settings for microsatellite analysis.

Results

MLVA-8 using capillary electrophoresis

The PCR primers described by Pourcel et al (2007) were used to amplify eight *Legionella pneumophila* specific VNTR markers (loci), named Lpms1b, 3, 13, 17, 19b, 33, 34 and 35 (Lpms stands for *Legionella pneumophila* minimicrosatellite, Pourcel et al., 2007). As can be seen in Table I, for some of the markers, the maximum size that could be expected based on data from Pourcel et al (2007) exceeded 600 bp, a size that is incompatible with the commercially available Applied Biosystems GeneScan LIZ size standards, as the largest fragments for these size standards are either 500 or 600 bp. However, we were able to use the newly developed Applied Biosystems GeneScan 1200 LIZ size standard, containing fragments up to 1200 bp, which allowed us to successfully determine the size of PCR products up to at least 932 bp.

Preliminary experiments showed that under regular PCR conditions, amplification using a number of the primer sets resulted in a series of peaks on the electropherogram, with one repeat size difference in length (not shown). This indicated possible binding of one of the primers inside the repeat region, which was confirmed 'in silico' for some of the primers (not shown). In order to avoid this effect, Touchdown PCR (Don et al., 1991) was used. This dramatically reduced the number of peaks, resulting in almost all cases to one sharp peak during the capillary electrophoresis.

Figure 1 shows two representative electropherograms for strain EUL 155, one for panel I (markers Lpms13, 17, 19b and 34) and one for panel II (markers Lpms1b, 3, 33 and 35). In each case, four differently colored peaks, each corresponding to a PCR product specific for one marker, can be seen.

Proficiency panel genotypes

The genotypes of the ten strains of the proficiency panel and the three reference strains used by Pourcel et al. (2007) were determined using capillary electrophoresis for PCR fragment size estimation. Alleles, as represented by fragment lengths observed, are shown in Table II. Based on PCR product size, the reported number of repeats and their repeat unit length for the reference strain Philadelphia-1 (Pourcel, et al., 2007), the total number of flanking bases were calculated, see Table I. These numbers were used to calculate the number of repeats for each marker based on the allele size (Pourcel, et al., 2007, see Table II). Table II also shows the number of repeats as determined by gel electrophoresis by Pourcel et al. (2007). There was an excellent correspondence between the number of repeats determined by both methods, with the exception of two alleles of marker Lpms1b. For this marker, allele 524 (strains EUL 048 and 056), corresponding to 7.5 repeats in the original MLVA-8 assay (Pourcel, et al., 2007), was predicted to have 7 (7.1) repeats in our experiments. Allele 611 (strain EUL 156), corresponding to 9.5 repeats in the original MLVA-8, was predicted to have 9 repeats. To obtain the actual length, the sequences of these fragments were determined. The results obtained indicated that the size of the fragments corresponding to allele 524 (strains EUL 048 and 056) was in fact 535 bp, while the size of the fragment corresponding to allele 611 (strain EUL 156) was 625 bp. These numbers are consistent with 7.5 (7.3) and 9.5 (9.3) repeats, respectively.

Duplicate or triplicate typings were performed to assess repeatability, see Table III. Usually, less than one basepair variation in fragment sizes between runs, or within the same run was observed, although a few alleles showed slightly more variation. The exception was marker Lpms34, which showed considerably more variation (Table III). However, due to the large repeat unit for his marker (125 bp), and the corresponding large distance between alleles (fragments), alleles could be identified accurately also for this marker.

Pourcel et al. (2007) reported no amplification for marker Lpms19b for two strains, EUL 154 and 155, a result which was consistent with reports from three institutes that performed typing using identical methods. As can be seen in Table II, in the present study, PCR fragments for this marker were obtained for both these strains. Sequencing these PCR products showed that they consisted of flanks and repeats that were similar to the Lpms19b fragments of the reference strains.

The observed allele sizes (by capillary electrophoresis) were compared with the sizes reported in the "Help File for the *Legionella pneumophila* MLVA typing page" (http://bacterial-genotyping.igmors.u-psud.fr/Legionella2006/help.htm), see Table IV. This help file is meant to assist in allele assignment for the gel-based MLVA genotyping assay. From the data in Table IV, it can be seen that the size of each allele as determined by CE differs, being usually more, but sometimes less than the theoretical size. The maximum difference observed was 15 bp (Lpms1b, CE allele 611). This difference in actual size and size observed using capillary electrophoresis has been shown for a *Bacillus anthracis* MLVA assay (Lista, et al., 2006). These authors showed up to 13 bp differences in predicted and observed sizes. As for *B. anthracis* (Lista, et al., 2006), the size differences observe in the present work increased with the number of repeats (fragment length, Table IV), except for marker Lpms34. The size differences were consistent and reproducible, and did not interfere with correct calling of alleles for most markers, except for marker Lpms1b (see above).

For marker Lpms33, it was possible to increase the number of alleles found in the strains of the proficiency panel from 4 to 6. Fragments corresponding to two repeats showed peaks that could consistently be resolved into two alleles after CE analysis, i.e. 348 and 349 bp (Table II and III). Strain EUL 25 showed the longer allele while strains Lens and EUL 156 showed the shorter allele. Also for fragments corresponding to four repeats, two alleles could be discerned, i.e. 597 bp (strain EUL 137) and 599 bp (strains Paris and EUL 157); the difference between the average fragment sizes for these alleles

was 1.69 bp (Table III). Sequencing of the corresponding PCR products confirmed that these fragments were in fact of different lengths. The following sizes were found for the Lpms1b PCR products: strains Lens and EUL 156: 352 bp; strain EUL 25: 353 bp; strain EUL 137: 603 bp; strains EUL 157 and Paris: 604 bp. For the 348 and 349 bp alleles, the corresponding PCR fragments did in fact differ by 1 bp in length. For the 597 and 599 bp alleles, the one bp size difference between the fragments resulted in a 1.69 bp difference in CE fragment sizes (Table III). Possibly, sequence-specific migration behavior influenced the different CE migration behaviors of the particular PCR fragments.

Fragments that differ only slightly in size can be clustered into one allele by the GeneMapper software if desirable.

Discussion

Multi-locus Variable number of Repeats Analysis (MLVA) allows for fast, easy, reproducible and portable genotyping of pathogenic microorganisms (Lindstedt, 2005, van Belkum, 2007). Fragment sizing using gel electrophoresis is an easy method to perform genotyping, however, it is difficult to obtain consistent running conditions across and between gels. Also, fragment size determination using agarose gels suffers from low resolution. As an alternative, capillary electrophoresis (CE) of DNA fragments can be used (van Belkum, 2007). Fragment sizing using CE can increase throughput, accuracy, resolution (up to a fraction of a basepair) and sensitivity. In the present work, an existing, agarose gel electrophoresis based MLVA scheme for *Legionella pneumophila* has been adapted to capillary electrophoresis. The method was applied to the ten strains of the so-called proficiency panel and the three reference strains (Philadelphia-1, Paris and Lens) that have been genotyped earlier using the same markers, but with gel-based fragment sizing (Pourcel, et al., 2007). The genotypes determined by our multiplexed assay corresponded very well with the ones obtained by Pourcel et al. (2007).

In two cases it was possible to optionally split an allele into two, slightly different sized alleles, which were confirmed by sequencing of the corresponding PCR products (marker Lpms33). This illustrates the greater sensitivity of capillary over gel electrophoresis.

Whereas Pourcel et al. (2007) failed to amplify marker Lpms19b in two strains (EUL 154 and 155), here PCR products corresponding to 4 repeats for these strains were observed, a result that was confirmed by sequencing the PCR products. It is likely that this difference in amplification success can be explained by the differences in the specific PCR conditions (choice of enzyme, cycling conditions) used.

In our assay, a new size standard from Applied Biosystems (GeneScan 1200 LIZ) was used, which in our hands successfully allowed the sizing of fluorescently labeled PCR

products up to almost 1000 bp. This size standard has a high density of fragments (at least every 20 bp) and thereby overcomes some of the problems reported with the GeneScan 500 LIZ as mentioned in the product insert (http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042491.pdf). To our knowledge, this work is the first paper describing the use of this size standard.

As shown for other capillary-based MLVA typing schemes (Lista, et al., 2006), the sizes determined in our assay deviated from the reported sizes for the original, gel-based MLVA-8. This inconsistency might be due to a combination of factors, such as a contribution of the different fluorescent dyes to migration distance, or sequence-specific migration behavior. It was noticed that using a different size ladder also results in different observed fragment sizes (our unpublished observations), an observation that is in line with results from investigations of the effect of size ladders on CE fragment size estimates (Akbari et al., 2007). Additionally, it is known that different instruments or polymers also have an effect on the fragment size results (Lindstedt et al., 2004). However, the differences between sizes observed using the present assay and theoretical fragment sizes are consistent and reproducible, and do not interfere with genotyping - with one exception. For marker Lpms1b, two of the alleles observed with our CE assay corresponded with repeats numbers that were a half-repeat shorter than predicted in the original, gel-based MLVA-8 assay. This difference is probably due to the relatively large size discrepancy between observed and reported sizes for this marker (from 11 bp up to 15 bp). With a repeat size of 45 bp for this marker, these differences are large enough to have an effect on the predicted number of repeats. Based on this observation it can be suggested that in general, VNTR markers whose migration deviates significantly from the expected migration pattern, should be avoided when designing or optimizing an MLVA assay using capillary electrophoresis.

On the other hand, as suggested elsewhere (Lista, et al., 2006), for any CE based MLVA scheme, a table which shows the correspondence between the sizes observed and the corresponding alleles needs to be provided. For our assay, Table IV serves this purpose.

The genotyping system employed here allows for fast, easy and automated genotyping of *L. pneumophila* strains. The genotypes obtained are compatible with those obtained by the gel electrophoresis assay and can be queried on the available *Legionella pneumophila* MLVA database (http://bacterial-genotyping.igmors.u-psud.fr/). In addition, this assay allows for an increased number of alleles in comparison with the gel-based genotyping assay, a result of the greater sensitivity of capillary electrophoresis.

Acknowledgments

The authors would like to thank Wenche Fonahn at the Norwegian National Institute for Public Health for performing the DNA isolations, the European Working Group on Legionella Infections, in particular Norman Fry, for providing the EUL strains and Dominique Caugant.

References

Akbari, A., Marthinsen, G., Lifjeld, J.T., Albregtsen, F., Wennerberg, L., Stenseth, N.C., Jakobsen, K.S., 2007. Improved DNA fragment length estimation in capillary electrophoresis. Electrophoresis. in press.

Borella, P., Guerrieri, E., Marchesi, I., Bondi, M., Messi, P., 2005. Water ecology of *Legionella* and protozoan: environmental and public health perspectives. Biotechnol Annu Rev. 11, 355-380.

Declerck, P., Behets, J., van Hoef, V., Ollevier, F., 2007. Detection of *Legionella* spp. and some of their amoeba hosts in floating biofilms from anthropogenic and natural aquatic environments. Water Res. 41, 3159-3167.

Den Boer, J.W., Yzerman, E.P., Schellekens, J., Lettinga, K.D., Boshuizen, H.C., Van Steenbergen, J.E., Bosman, A., Van den Hof, S., Van Vliet, H.A., Peeters, M.F., Van Ketel, R.J., Speelman, P., Kool, J.L., Conyn-Van Spaendonck, M.A., 2002. A large outbreak of Legionnaires' disease at a flower show, the Netherlands, 1999. Emerg Infect Dis. 8, 37-43.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., Mattick, J.S., 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19, 4008.

Gaia, V., Fry, N.K., Afshar, B., Luck, P.C., Meugnier, H., Etienne, J., Peduzzi, R., Harrison, T.G., 2005. Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. J Clin Microbiol. 43, 2047-2052.

Henry, B., Young, J.G., Walker, D.M.C., 2005. Report of the expert panel on the Legionnaires Disease outbreak December 2005, Ontario Ministry of Health and Long-Term Care, Ontario, USA.

Lindstedt, B.A., 2005. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. Electrophoresis. 26, 2567-2582.

Lindstedt, B.A., Vardund, T., Kapperud, G., 2004. Multiple-Locus Variable-Number Tandem-Repeats Analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. J Microbiol Methods. 58, 213-222.

Lista, F., Faggioni, G., Valjevac, S., Ciammaruconi, A., Vaissaire, J., le Doujet, C., Gorge, O., De Santis, R., Carattoli, A., Ciervo, A., Fasanella, A., Orsini, F., D'Amelio, R., Pourcel, C., Cassone, A., Vergnaud, G., 2006. Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC Microbiol. 6, 33.

McDade, J.E., 2002. Legionnaires' disease 25 years later: Lessons learned. In: R. Marre, Y. A. Kwaik, C. Bartlett, N. P. Cianciotto, B. S. Fields, M. Frosch, J. Hacker, P. C. Lück (Eds.), Legionella, ASM Press, Washington, pp. 1-10.

Nygard, K., 2005. Update: outbreak of legionnaires disease in Norway traced to air scrubber. Euro Surveill. 10, E050609 050601.

Pourcel, C., Vidgop, Y., Ramisse, F., Vergnaud, G., Tram, C., 2003. Characterization of a tandem repeat polymorphism in *Legionella pneumophila* and its use for genotyping. J Clin Microbiol. 41, 1819-1826.

Pourcel, C., Visca, P., Afshar, B., D'Arezzo, S., Vergnaud, G., Fry, N.K., 2007. Identification of variable-number tandem-repeat (VNTR) sequences in *Legionella pneumophila* and development of an optimized multiple-locus VNTR analysis typing scheme. J Clin Microbiol. 45, 1190-1199.

Ratzow, S., Gaia, V., Helbig, J.H., Fry, N.K., Luck, P.C., 2007. Addition of neuA, the gene encoding N-acylneuraminate cytidylyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing *Legionella pneumophila* serogroup 1 strains. J Clin Microbiol. 45, 1965-1968.

van Belkum, A., 2007. Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). FEMS Immunol Med Microbiol. 49, 22-27.

van Soolingen, D., de Haas, P.E.W., Kremer, K., 1999. Restriction fragment length polymorphism (RFLP) typing of *Mycobacteria*, National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands.

Yokoyama, E., Kishida, K., Uchimura, M., Ichinohe, S., 2006. Comparison between agarose gel electrophoresis and capillary electrophoresis for variable numbers of tandem repeat typing of *Mycobacterium tuberculosis*. J Microbiol Methods. 65, 425-431.

Table I: Setup of the *Legionella pneumophila* MLVA-8 using capillary electrophoresis. Minimum and maximum lengths based on the minimum and maximum number of repeats reported for the proficiency panel, including the three reference strains, in Pourcel et al. (2007)

Primer set	Dye	Color[a]	Panel	Repeat length in bp	Total flank in bp[b]	Expected length in bp	
						minimum	maximum
Lpms1b	VIC	Green	II	45	205	475	633
Lpms3	FAM	Blue	II	96	173	845	941
Lpms13	FAM	Blue	I	24	164	260	548
Lpms17	PET	Red	I	39	200	259	278
Lpms19b	NED	Yellow	I	21	89	173	194
Lpms33	NED	Yellow	II	125	102	227	602
Lpms34	VIC	Green	I	125	84	209	459
Lpms35	PET	Red	II	18	148	202	580

[a] The GeneMapper software 'translates' the fluorescent dyes into these colors

[b] Deduced from the PCR product sizes and number of repeats for strain Philadelphia-1 according to Pourcel et al. (2007), table 3

Table II. Comparison of MLVA-8 based on capillary electrophoresis (CE), and based on gel electrophoresis. CE Alleles in bp, Gel repeats: number of repeats based on gel electrophoresis as described in Pourcel et al. (2007). NA: no amplicon reported.

Marker	Lpms1b			Lpms3		
Strain	CE allele	CE repeats	Gel repeats[a]	CE allele	CE repeats	Gel repeats[a]
Phil_NCTC11 92	553	7.7	8	932	7.9	8
Paris_EUL 146	509	6.8	7	837	6.9	7
Lens_EUL 160	465	5.8	6	837	6.9	7
EUL 025	597	8.7	9	932	7.9	8
EUL 048	524	7.1	7.5	837	6.9	7
EUL 056	524	7.1	7.5	837	6.9	7
EUL 121	509	6.8	7	932	7.9	8
EUL 137	597	8.7	9	932	7.9	8
EUL 153	553	7.7	8	932	7.9	8
EUL 154	553	7.7	8	932	7.9	8
EUL 155	553	7.7	8	932	7.9	8
EUL 156	611	9.0	9.5	932	7.9	8
EUL 157	509	6.8	7	837	6.9	7

Marker	Lpms13			Lpms17		
Strain	CE allele	CE repeats	Gel repeats[a]	CE allele	CE repeats	Gel repeats[a]
Phil_NCTC11 92	429	11.0	11	282	2.1	2
Paris_EUL 146	405	10.0	10	282	2.1	2
Lens_EUL 160	259	4.0	4	282	2.1	2
EUL 025	551	16.1	16	282	2.1	2
EUL 048	454	12.1	12	282	2.1	2
EUL 056	454	12.1	12	282	2.1	2
EUL 121	381	9.0	9	282	2.1	2
EUL 137	381	9.0	9	282	2.1	2
EUL 153	429	11.0	11	257	1.5	1.5
EUL 154	429	11.0	11	257	1.5	1.5
EUL 155	429	11.0	11	257	1.5	1.5
EUL 156	357	8.0	8	282	2.1	2
EUL 157	405	10.0	10	282	2.1	2

Marker	Lpms19b			Lpms33		
Strain	CE allele	CE repeats	Gel repeats[a]	CE allele	CE repeats	Gel repeats[a]
Phil_NCTC11 92	170	3.9	4	224	1.0	1
Paris_EUL 146	170	3.9	4	599[c]	4.0	4
Lens_EUL 160	170	3.9	4	348[b]	2.0	2
EUL 025	170	3.9	4	349[b]	2.0	2
EUL 048	170	3.9	4	473	3.0	3
EUL 056	170	3.9	4	473	3.0	3

EUL 121	191	4.9	5	473	3.0	3
EUL 137	170	3.9	4	597 [c]	4.0	4
EUL 153	170	3.9	4	224	1.0	1
EUL 154	170	3.9	NA	224	1.0	1
EUL 155	170	3.9	NA	224	1.0	1
EUL 156	170	3.9	4	348 [b]	2.0	2
EUL 157	170	3.9	4	599 [c]	4.0	4

Table II, continued

Marker	Lpms34			Lpms35		
Strain	CE allele	CE repeats	Gel repeats[a]	CE allele	CE repeats	Gel repeats[a]
Phil_NCTC11 92	205	1.0	1	201	2.9	3
Paris_EUL 146	331	2.0	2	449	16.7	17
Lens_EUL 160	454	3.0	3	555	22.6	23
EUL 025	331	2.0	2	432	15.8	16
EUL 048	205	1.0	1	449	16.7	17
EUL 056	205	1.0	1	449	16.7	17
EUL 121	205	1.0	1	361	11.8	12
EUL 137	205	1.0	1	573	23.6	24
EUL 153	205	1.0	1	201	2.9	3
EUL 154	205	1.0	1	201	2.9	3
EUL 155	205	1.0	1	201	2.9	3
EUL 156	331	2.0	2	467	17.7	18
EUL 157	331	2.0	2	449	16.7	17

[a] Taken from Pourcel et al. (2007)

[b] Two alleles found for CE, only one with gel electrophoresis

[c] Two alleles found for CE, only one with gel electrophoresis

Table III: Allele statistics. Fragment sizes obtained from duplicate or triplicate typings were pooled for each allele across all strains. Average: average size determined by CE for each allele ± 1 standard deviation. N: Number of fragment sizes acquired for the allele

Lpms1b			Lpms3			Lpms13		
Allele	**Average**	**N**	**Allele**	**Average**	**N**	**Allele**	**Average**	**N**
465	465.44 ± 0.17	7	837	836.94 ± 0.26	34	259	259.45 ± 0.03	8
509	509.14 ± 0.15	20	932	931.82 ± 0.23	55	357	356.70 ± 0.03	6
524	523.90 ± 0.18	14				381	380.89 ± 0.08	10
553	552.71 ± 0.19	27				405	404.50 ± 0.04	12
597	596.89 ± 0.44	15				429	429.33 ± 0.04	23
611	610.93 ± 0.19	9				454	453.53 ± 0.04	12
						551	551.22 ± 0.05	6

Lpms17			Lpms19			Lpms33		
Allele	**Average**	**N**	**Allele**	**Average**	**N**	**Allele**	**Average**	**N**
257	257.22 ± 0.41	18	170	169.53 ± 0.15	67	224	223.68 ± 0.07	27
282	281.98 ± 0.43	54	191	190.53 ± 0.05	5	348	348.20 ± 0.16	16
						349	349.26 ± 0.12	7
						473	472.55 ± 0.23	20
						597	597.40 ± 0.14	8
						599	599.09 ± 0.10	12

Lpms34			Lpms35		
Allele	**Average**	**N**	**Allele**	**Average**	**N**
205	205.38 ± 0.57	50	201	200.86 ± 0.06	16
331	330.52 ± 2.19	24	361	360.90 ± 0.10	4
454	453.94 ± 1.56	6	432	432.00 ± 0.10	4
			449	449.14 ± 0.24	16
			467	467.12 ± 0.21	6
			555	555.30 ± 0.30	4
			573	573.17 ± 0.18	4

Table IV: Comparison of PCR product length inferred from capillary electrophoresis and the sizes from the *L. pneumophila* MLVA-8 allele assignment table

Lpms1b				Lpms3				Lpms13				Lpms17			
Rep[a]	Size[b]	CE[c]	Diff[d]	Rep	Size	CE	Diff	Rep	Size	CE	Diff	Rep	Size	CE	Diff
6	475	465	-10	7	846	837	-9	4	260	259	-1	1.5	258	257	-1
7	520	509	-11	8	941	932	-9	8	356	357	1	2	278	282	4
7.5	535	524	-11					9	380	381	1				
8	565	553	-12					10	404	405	1				
9	610	597	-13					11	428	429	1				
9.5	626	611	-15					12	452	454	2				
								16	548	551	3				

Lpms19b				Lpms33				Lpms34				Lpms35			
Rep	Size	CE	Diff	Rep	Size	CE	Diff	Rep	Size	CE	Diff	Rep	Size	CE	Diff
4	173	170	-3	1	227	224	-3	1	209	205	-4	3	202	201	-1
5	194	191	-3	2	352	348	-4	2	334	331	-3	12	364	361	-3
						349	-3	3	460	454	-6	16	436	432	-4
				3	477	473	-4					17	454	449	-5
				4	604	597	-7					18	472	467	-5
						599	-5					23	562	555	-7
												24	580	573	-7

[a] Number of repeats based on the "Help File for the *Legionella pneumophila* MLVA typing page" (http://bacterial-genotyping.igmors.u-psud.fr/Legionella2006/help.htm)

[b] Size of the fragments based on the "Help File for the *Legionella pneumophila* MLVA typing page"

[c] Size of the alleles as determined by capillary electrophoresis in this study

[d] Diff: difference between the sizes of the previous two columns

Figures Legends

Figure 1: Electropherograms (screenshot of the GeneMapper software) showing separation of four fluorescently labeled PCR products for PCRs of strain EUL 155 of panel I markers (A) and panel II markers (B).